

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.

Dnr 01-2003:2038

FÖRORD

Föreliggande rapport har utarbetats vid enheten för resultatbedömning vid Skolverkets utredningsavdelning. Rapporten har tillkommit för att ge underlag för Skolverkets svar på ett regeringsuppdrag att göra en ”bedömning av dagens provsystem med avseende på kvalitet och kostnadseffektivitet”. Framför allt den inte helt lätta frågan om vad som i sammanhanget kan eller bör läggas in i begreppet kvalitet har lett till att rapporten blivit rätt omfattande. Utredningsgruppen gör trots detta givetvis inte anspråk på att ha täckt in alla aspekter, men hoppas att rapporten kan tjäna som underlag för en fortsatt diskussion om provsystemet och dess roll i hela skolsystemet.

Följande personer har ingått i den arbetsgrupp vid enheten som utarbetat underlaget: Jan Engstedt, Karin Hector-Stahre, Tommy Lagergren, Anna Lena Larsson, Trine Normann, Bo Palaszewski, Kristian Ramstedt och Eva Wirén. Värdefulla synpunkter har också lämnats av medarbetare vid andra enheter inom Skolverket, Anna Barklund, Cecilia Bergström, Anita Börlin, Bengt Fredén, Ulla Lindqvist och Jonas Sandqvist samt av Gudrun Erickson vid Göteborgs universitet och Astrid Pettersson vid Lärarhögskolan i Stockholm.

Bengt Fredén har skrivit delar av kapitel 1, Eva Wirén har i samverkan med Trine Normann skrivit kapitel 2. Kristian Ramstedt har lett arbetet och skrivit övriga delar.

Staffan Lundh

Avdelningschef

Kristian Ramstedt
Enhetschef

Dnr 01-2003:2038

INNEHÅLL

SAMMANFATTNING OCH FÖRSLAG TILL SKOLVERKETS

BEDÖMNING.....9

PROVSYSTEMETS OMFATTNING, TEKNISKA KVALITET OCH KOSTNADSEFFEKTIVITET	9
<i>Provens tekniska kvalitet</i>	10
<i>Provsystemet kostnadseffektivitet</i>	10
Utvecklingskostnader	10
Övriga kostnader	11
Sekretessen över tid.....	12
BREDDNING AV DAGENS PROVSYSTEM TILL NYA ÄMNEN OCH ÅLDRAR.....	13
PROVSYSTEMETS KVALITET UR ETT SYSTEMPERSPEKTIV	14
Provsystemet som stöd för en likvärdig och rättvis betygssättning.....	14
Provsystemet ska konkretisera kursmål och betygskriterier ..	16
Provsystemets diagnostiska funktion	16
Provsystemet ska ge underlag för uppföljning.....	17
Proven styr innehåll och arbetsmetoder	17
Proven används som examensprov	18
VAD, VARFÖR OCH VARTHÄN?	18
<i>Alternativ 1: Ett nationellt provsystem med tonvikt på läroplanens intentioner</i>	19
<i>Alternativ 2: Ett nationellt provsystem för instrumentell kalibrering av betyg och generering av uppföljningsdata.</i>	21
<i>Överväganden</i>	22
UTKAST TILL NYTT PROVSYSTEM	23
<i>Grundskolan och motsvarande skolformer</i>	23
<i>Gymnasieskolan och motsvarande skolformer</i>	24
<i>Gemensamt förslag för alla skolformer</i>	25
<i>Förslagets koppling till det nationella provsystemets syften</i>	26
<i>Kostnad</i>	26
<i>Provbanken</i>	27

INLEDNING29

1. UTFORMNING, OMFATTNING OCH KOSTNADER.....31

<i>Tidigare provuppdrag</i>	31
<i>Uppdrag 1994</i>	31
Grundskolan	31
Den gymnasiala utbildningen.....	33
Provbank och nationella utvärderingar	33
<i>Uppdrag 1999</i>	34
Bakgrund	34
Grundskolan	35
Den gymnasiala utbildningen.....	36
Provbanken.....	36

Dnr 01-2003:2038

<i>Sammanfattande kommentar</i>	37
PROVVERKSAMHETEN LÄSÅRET 2002/2003	37
<i>Grundskolan</i>	37
<i>Gymnasieskolan</i>	39
<i>Vuxenutbildningen och övriga skolformer</i>	42
VILKA KONSTRUERAR PROVEN?	42
KOSTNAD, SÅRBARHET OCH SEKRETESS	43
<i>Provs kvalitet</i>	44
<i>Hur lång tid behövs för att utveckla prov?</i>	45
<i>Kostnader för provkonstruktion</i>	45
USA	46
Kanada	47
Nederländerna	47
Norden	48
<i>Sverige</i>	49
Svenska	49
Engelska	49
Matematik	50
Svenska för invandrare, Sfi	50
Provbanksprov	50
Distribution och tryck	51
<i>Slutsatser</i>	52
DE NATIONELLA PROVEN SÅRBARHET	52
<i>Grundskolan</i>	53
<i>Gymnasieskolan</i>	54
<i>Rättsläget</i>	55
<i>Förslag till åtgärder för att minska sårbarheten</i>	55
HUR LÄNGE SKA PROVEN VARA HEMLIGA?	56
<i>Produktkvalitet och processkvalitet</i>	57
<i>Varför sekretess?</i>	57
<i>Konsekvenser av sekretessen</i>	58

2. VAD ANSER ANVÄNDARNA OM PROVSYSTEMET?60

INLEDNING	60
OM DET NUVARANDE PROVSYSTEMETS KVALITET – ERFARENHET OCH VÄRDERING FRÅN SKOLFÄLTET	61
<i>Vad anser rektorer och lärare om de nationella proven och det nuvarande nationella provsystemet?</i>	61
Proven möter skolan	61
<i>Om provens (olika) syften och relativa värde</i>	62
Uppfyller proven sina syften?	62
Ses proven som viktiga eller oviktiga?	64
<i>Hur påverkar proven skolans arbete och undervisning och hur stämmer de med lokala mål och kriterier?</i>	66
Påverkar/styr proven undervisningen?	66
Stämmer proven överens med lokala mål?	68
<i>Hur används de nationella diagnostiska materialen, om de överhuvudtaget används?</i>	70

Dnr 01-2003:2038

<i>Hur använder skolan sig av provresultaten - förmedling, samverkan, samarbete?</i>	72
<i>Vad säger lärarna om elevers reaktioner inför proven och vad säger eleverna om proven?</i>	73
<i>Lärares och rektorers uppfattning om proven och provsystemet – några övergripande attitydfrågor i enkäten</i>	76
NÅGRA HUVUDSAKLIGA RESULTAT OCH PRELIMINÄRA SLUTSATSER	
– EN SAMMANFATTNING	79

3. NÅGRA JÄMFÖRELSE AV PROVBETYG OCH SLUTBETYG82

GRUNDSKOLAN	83
<i>Jämförelser mellan ämnen</i>	84
<i>Jämförelser mellan skolor</i>	85
Svenska	85
Engelska	86
Matematik	87
GYMNASIESKOLAN	88
<i>Jämförelse mellan ämnen</i>	89
<i>Jämförelser mellan skolor och program</i>	90
Svenska	90
Matematik	94
DISKUSSION	96
<i>Är provbetyg och slutbetyg jämförbara?</i>	97
Provbetygen	97
Slutbetyget i ämnet eller kursen	98
Relationen provbetyg - slutbetyg	98
<i>Skillnader mellan ämnen</i>	102
Poängjämförelse av matematikprov och ett delprov i engelska	103
Har ämnena olika karaktär?	105
<i>Skillnader mellan olika skolor</i>	107
Skolstorlekens betydelse	107
Samband mellan skolmedelvärden för provbetyg och slutbetyg	107
Finns ”högprovgsskolor” i urvalet?	108
<i>Skillnader mellan olika program</i>	110
<i>Skillnader mellan grundskola och gymnasieskola</i>	110
RELATIONEN MELLAN PROVBETYG OCH SLUTBETYG GENERELLT	112
<i>Prov som mått på den enskilda elevens måluppfyllelse</i>	112
<i>Prov som mått på den enskilda gruppens måluppfyllelse</i>	113
<i>Prov som förebildligt stödmaterial</i>	114

4. LÄROPLANER OCH PROVSYSYSTEM.....117

OLIKA TYPER AV LÄROPLANER	117
1. Läroplanen som ett formellt reglerande dokument	117
2. Läroplanen som ett auktoriserande och innehållsbeskrivande dokument	118

Dnr 01-2003:2038

3. Läroplanen som ett politiskt normerande instrument	118
4. Läroplanen som ett standardiserande dokument för central utvärdering (evaluering) av utbildningen	119
VILKEN LÄROPLANSTYP HAR SVERIGE IDAG?	120
SVENSK KUNSKAPSBEDÖMNING UNDER OLIKA LÄROPLANER	120
BETYGSFRÅGAN	123
Värdegrundsmål och betyg	125
Likvärdigheten.....	126
5. NATIONELLA PROV I FLERA ÄMNER OCH FÖR FLERA ÅLDRAR?.....	129
PROV I FLERA ÄMNER OCH FÖR FLERA ÅLDRAR?	129
Vill lärare ha prov i flera ämnen?	129
Enkätstudie i grundskolan	129
Hearingen	130
Vad anser eleverna om prov i flera ämnen?	131
Effektstudien	131
Hearingen	132
Slutsatser	132
BREDDNING MED HJÄLP AV ANDRA INSTRUMENT ÄN NATIONELLA PROV	133
Internationella studier.....	133
Nationell kunskapsutvärdering	134
Sammanfattning.....	135
6. NÅGRA AVSLUTANDE SYNPUNKTER, SAMMANFATTNING OCH FÖRSLAG TILL STÄLLNINGSTAGANDEN	137
SIGNALER.....	137
KVALITET.....	137
LIKVÄRDIGHET OCH RÄTTVISA.....	138
Likvärdighet	139
Rättvisa.....	140
SLUTSATSER.....	140
SAMMANFATTNING OCH FÖRSLAG.....	143
 APPENDIX	
KUNSKAPSBEDÖMNING OCH PROV – EN ÖVERSIKT.....	144
INLEDNING	144
NÅGRA CENTRALA BEGREPP.....	145
STYRDOKUMENT OCH STANDARDS	149
Standards.....	150
Exemplet matematik.....	151
Olika nivåer	152
Instrument som baseras på standards.....	152
National Assessment of Educational Progress (NAEP).....	152
PISA	153

Dnr 01-2003:2038

<i>Standards, betyg och likvärdig bedömning</i>	154
ÖPPNA ELLER SLUTNA STYRDOKUMENT?	155
PROV	156
<i>Validitet och reliabilitet</i>	158
<i>Klassisk och modern testteori</i>	159
<i>Normrelaterade prov</i>	160
<i>Kriterierelaterade prov</i>	161
<i>Standardrelaterade prov</i>	162
<i>Standard setting</i>	163
Angoff's metod.....	164
Contrasting groups	165
Bookmarking.....	166
<i>Några ytterligare aspekter</i>	167
<i>Vad används provresultat till?</i>	168
ÄR ALLA NÖJDA MED TÄNKANDET I STANDARDS?	169
REFERENSER	171

Dnr 01-2003:2038

SAMMANFATTNING OCH FÖRSLAG TILL SKOLVERKETS BEDÖMNING

Utredningsgruppen sammanfattar här sina ställningstaganden samt ger förslag till bedömningar från Skolverkets sida.

Uppdrag E lyder i sin helhet

Skolverket bör redovisa verkets bedömning av dagens system med nationella prov avseende kvalitet och kostnadseffektivitet. Verket bör analysera försättningsarna för att bredda systemet med nationella ämnesprov till dels fler åldersgrupper i grundskolan, dels fler ämnen eller ämnesgrupper. Utgångspunkt bör vara både den lokala nivåns och den nationella nivåns behov av uppföljning. Detta bör ske såväl för grundskolans och motsvarande skolformer som för gymnasieskolan och motsvarande skolformer.

Uppdraget innebär att Skolverket skall ge sin syn på provsystemets kvalitet, dess kostnadseffektivitet, möjligheterna till breddning av systemet samt i vilken utsträckning provsystemet kan fylla behov av uppföljning.

Sammanfattningen tar först upp de delar som utredningsgruppen finner tämligen enkla att hantera. Det gäller bedömningen av provens tekniska kvalitet, kostnadseffektivitet och breddning till fler ämnen och åldrar. Därefter fokuseras de svårare bedömningar som har att göra med provsystemets olika syften och effekter, i vilken utsträckning dessa är i enlighet med skolsystemet och hur förenliga de är inom ramen för ett enda provsystem. Ett par alternativa utformningar som kan ses som ytterligheter utifrån olika syften presenteras med för- och nackdelar. Enligt utredningsgruppens bedömning är dock inget av dessa renodlade alternativ acceptabelt och därför lämnas ett förslag till förändrat provsystem som enligt utredningsgruppens bedömning är en acceptabel kompromiss.

Provsystemets omfattning, tekniska kvalitet och kostnadseffektivitet

I kapitel 1 i utredningen redogörs för det nuvarande provsystemets omfattning. Det baseras på regeringens uppdrag från 1999 där det fastläggs att ämnesproven i år 9, kärnämnesproven samt proven i de avslutande gemensamma karaktärsämneskurserna skall vara obligatoriska. Vidare anges att det skall finnas en provbank som skall innehålla prov, delar av prov och diagnostiska material för i första hand ämnen och kurser i vilka det inte finns nationella prov.

Uppdraget innebär för grundskolans del **6 årliga nationella prov** och **5 diagnostiska material**. Tre obligatoriska prov tillhandahålls i år 9 och tre frivilliga i år 5. De diagnostiska materialen gäller åren F - 5

Dnr 01-2003:2038

respektive 6 – 9, i det senare fallet i alla tre provämnerna, i det förra endast i svenska/svenska som andraspråk och matematik.

I gymnasieskolan erbjuds årligen **14 obligatoriska nationella prov**, 8 i matematik, 4 i engelska och 2 i svenska. Via provbank erbjuds för närvarande **11 frivilliga prov** i franska, tyska, matematik, fysik och biologi. Vidare pågår utvecklingsarbetet med ytterligare **16 provbanksprov** i olika ämnen eller för olika program.

För Sfi (svenska för invandrare) konstrueras årligen **2 nationella prov**, dessutom finns gamla prov tillgängliga i provbanken. Under 2003 utarbetas dessutom ett särskilt bedömningsmaterial för Sfi.

Sammantaget tillhandahålls således årligen 33 färdigutvecklade prov, dessutom pågår utvecklingsarbete inom provbankens ram för ytterligare 15. Sammantaget gör detta ett provprogram som innefattar **49 prov, 5 diagnostiska material och 1 bedömningsmaterial**.

Proven konstrueras vid ett antal universitetsinstitutioner på Skolverkets uppdrag. Skolverkets sammanlagda kostnad för provverksamheten under 2003 uppgår till knappt 45 miljoner kronor.

Provens tekniska kvalitet

Skolverket har under lång tid följt de olika provinstitutionernas arbete och har god insyn i konstruktionsarbetet.

Utredningsgruppen bedömer att konstruktionen av de svenska nationella proven följer internationell standard för professionell konstruktion av prov och att de därmed är av hög teknisk och innehållslig kvalitet.

Provsystemet kostnadseffektivitet

Provsystemet är omfattande och det är svårt att ange någon mer detaljerad bedömning av kostnadseffektiviteten. Varje institution konstruerar flera olika prov och beställningarna är inte specificerade på ett sådant sätt att kostnaden för varje enskilt prov kan urskiljas. I Skolverkets kostnader för provsystemet ingår inte enbart utvecklingskostnader utan även kostnader för tryckning och distribution av prov samt insamling av resultat.

Utvecklingskostnader

Ett sätt att försöka bedöma provens kostnadseffektivitet är att jämföra med motsvarande kostnader i andra länder. Det är inte lätt att få fram sådana uppgifter och ofta räknas olika poster in i kostnaden i olika länder. De siffror som anges här får därmed ses som tämligen grova skattningar. De kostnader som presenteras gäller själva utvecklingen av prov från det att konstruktionen av uppgifter börjar tills ett färdigt prov föreligger, en cykel på 1,5 till 2 år. Kostnader för tryckning, distribution och rättning ingår inte i de redovisade summorna.

Dnr 01-2003:2038

Uppgifterna gäller genomsnittlig kostnad för att utveckla ett prov och uttrycks i miljoner kronor.

Sverige

Svenska	1,2	miljoner kronor
Engelska	1,3	
Matematik	1,0 – 1,3	
SFI	0,85	

Andra länder

Danmark	0,1-0,2	miljoner kronor
Finland	1,1	
Norge	1,4	
USA	0,2 – 1,7	
Kanada	1,4	
Nederländerna	1,3	

Vid en jämförelse av detta slag framstår inte de svenska kostnaderna som avvikande eller anmärkningsvärda. Om man därtill betänker att ett grundskoleprov görs av cirka 100 000 elever blir kostnaden per elev cirka 12 kronor. I ljuset av att en grundskoleelev kostar 62 500 kronor per år framstår provkostnaden som rätt marginell.

Det sagda innebär naturligtvis inte att verksamheten inte kan effektiviseras, men några uppenbara effektivitetsvinster kan inte upptäckas utan en mer ingående analys av de interna kostnaderna vid provinstitutionerna.

Utredningsgruppens bedömning är att kostnaderna för utveckling av prov är rimliga i relation till de kvalitetskrav som gäller för framställningen.

Övriga kostnader

Det nuvarande provsystemet är sårbart. Det visades våren 2003 när någon offentliggjorde två delprov i matematik för år 9 på Internet innan proven skulle genomföras. Idag är det mycket enkelt att avbilda och publicera bilder på Internet och problemet att hålla proven hemliga är svårlöst.

Skolverkets kostnad för tryckning och distribution av prov uppgår för närvarande till cirka 3,5 miljoner kronor. Denna summa gäller i huvudsak grundskolan. Gymnasieskolan erhåller ett kopieringsunderlag, men får själv betala om färdigtryckta prov önskas.¹ Detta förfarande gör gymnasieproven sårbara. Om Skolverket även skulle betala tryckning och distribution av prov till gymnasieskolan uppskattas

¹ Detta är en besparingsåtgärd som infördes 2000.

Dnr 01-2003:2038

den sammanlagda kostnaden för grundskolan och gymnasieskolan uppgå till cirka 10,5 miljoner kronor årligen.

För närvarande distribueras proven utan särskild förslutning av den typ ("svarta påsar") som förekom när standardprov och centrala prov distribuerades. Om sådan distribution åter skulle införas skulle kostnaderna för provdistributionen öka väsentligt.

En skärpning av säkerhetshanteringen kring proven bidrar troligen också till att ytterligare dramatisera dem och ge dem karaktär av examensprov. Detta gör det i sin tur mer angeläget att försöka komma över proven i förtid.

Sårbarheten är också ett problem för totalinsamlingen av provresultat eftersom det blir svårt att få trovärdighet för resultat som gäller prov som varit publicerade i förtid.

Problemet med sårbarheten är svårlöst. Ett mer flexibelt och säkert distributionssystem måste skapas och sårbarheten måste minskas utan att stämningen och stressen kring proven haussas upp ytterligare. Den lösning på sikt som framstår som mest kostnadseffektiv och flexibel är elektronisk distribution. Det är dock tveksamt om skolornas utrustning idag håller sådan kvalitet att detta kan genomföras i stor skala.

Utredningsgruppen ser ett elektroniskt distributionssystem som framtidens lösning och föreslår att Skolverket får i uppdrag att utreda förutsättningarna för en säker och tillförlitlig elektronisk distribution av prov.

Sekretessen över tid

Sekretessfrågan gäller inte bara tiden fram till provet. De nationella proven är för närvarande sekretessbelagda i tio år (engelska och matematik). Denna sekretess har införts för att förbilliga provproduktionen och ge underlag för kunskapsuppföljning över tid.

Sekretessen har dock uppenbara negativa konsekvenser genom att prov och bedömningsmaterial inte kan bilda underlag för en professionell pedagogisk – didaktisk debatt. Det innebär också att det är svårt eller omöjligt för elever och föräldrar att få tillgång till proven för egna bedömningar av resultaten.

Återanvändningen av uppgifter varierar mellan olika provinstitutioner, för t.ex. engelska och vissa matematikprov innebär återanvändningen en inte oväsentlig besparing medan den för vissa matematikprov är närmast försumbar. Proven i svenska är inte sekretessbelagda och där sker ingen återanvändning.

Utredningsgruppen föreslår att den nuvarande tioåriga sekretessen avskaffas eller modifieras och att prov och bedömningsmaterial i ökad utsträckning blir tillgängliga när proven är genomförda.

Dnr 01-2003:2038

Breddning av dagens provsystem till nya ämnen och åldrar

I kapitel 2 i utredningen ges en preliminär redovisning av resultat från den effektstudie av provsystemet som genomförts i grundskolan våren 2003 och i kapitel 5 diskuteras frågan om breddning.

Generellt kan man konstatera att såväl lärare som rektorer i grundskolan är nöjda med det nuvarande provsystemet. De tycker att proven är arbets- och tidskrävande men att resultatet är värt den nedlagda arbetsinsatsen.

Även eleverna i grundskolan uttrycker i allmänhet att det är bra med nationella prov. De ser dem som en form av trygghet när de vet att alla andra gör samma prov.

Lärare och rektorer är således i hög grad positiva till nuvarande nationella prov och diagnostiska material. Men i nästan lika hög grad är de negativa till prov i nya ämnen och för nya åldrar. Det är dock svårt att veta vad de föreställer sig för typ av prov. Om de bedöms bli lika tidskrävande som nuvarande prov är det förståeligt om man rygggar tillbaka för ytterligare ett antal prov. I vilket fall kan man konstatera att någon önskan om fler prov är svår att spåra.

Något säkert underlag för att bedöma elevernas uppfattning i den här frågan finns inte, men vissa intervjudata kan tyda på att de är mer positiva än lärare och rektorer.

De lärare som svarat på enkäten har själva de nuvarande provämnena och det är möjligt att lärare, särskilt i år 9, som har andra ämnen skulle uttrycka en annan mening. Vid en hearing som genomfördes i september 2003 framkom dock inga tydliga signaler om att så är fallet, även om vissa representanter för samhällsvetenskapliga ämnen ställde sig positiva. Även bland representanter för praktisk-estetiska ämnen fanns förespråkare för prov, men även uttalade motståndare.

De två representanter för elevorganisationer som deltog önskade ingen breddning. De var uttalade motståndare till nationella prov. De ansåg att de i alltför hög grad styrde undervisningen och att de lade för stor vikt vid individutvärdering när det var skolan och skolans insatser som borde utvärderas. Dessutom sade de sig ha mycket varierande erfarenheter om dels vilken information de fått om de nationella provens roll, dels om hur resultaten använts.

Även lärare uttryckte uppfattningar om att proven riskerade att bli alltför styrande för undervisningen och hämmande för pedagogiskt nytänkande. Det fanns dock uppfattningar om att proven tvärtom hade en viktig funktion i det senare avseendet.

Från ett håll har det kommit tydliga signaler om att man önskar ett nationellt prov och det är kemisternas intresseorganisationer som uttryckt djup oro för kemiämnets situation i gymnasieskolan.

Dnr 01-2003:2038

Utredningsgruppen anser att det inte finns skäl att generellt utvidga nuvarande provsystem till fler ämnen och åldrar. För vissa ämnen kan det dock finnas särskilda skäl som motiverar en utvidgning.

Provsystemets kvalitet ur ett systemperspektiv

I kapitel 1 ges en kort resumé över provsystemets utveckling sedan de nu gällande läroplanerna infördes. Det nationella provsystemets syften har delvis varierat över tid och inte alltid uttryckts särskilt tydligt. I provsystemet inräknas de obligatoriska nationella proven, provbanksproven och de diagnostiska materialen. Utredningsgruppen tycker sig kunna identifiera följande syften.

- Provsystemet skall stödja en likvärdig och rättvis betygssättning (betygsstöd)
- Provsystemet skall konkretisera kursmål och betygskriterier (kommentarmaterial)
- Provsystemet skall visa på elevers starka och svaga sidor (diagnostisk funktion)
- Provsystemet skall genom insamling av resultat ge en bild av i vilken utsträckning kunskapsmålen nås (uppföljning).

Dessutom kan man spåra ett par effekter av de nationella proven som kan vara mer eller mindre önskvärda

- Proven styr innehåll och arbetsmetoder
- Proven används som examensprov.

De olika syftena är ibland svårförenliga och representerar helt eller delvis olika principiella ställningstaganden.

Provsystemet som stöd för en likvärdig och rättvis betygssättning

Detta är ett problematiskt syfte. Skolverket har aldrig varit särskilt tydligt med vad som avses med formuleringen. I kapitel 3 i utredningen beskrivs och diskuteras relationen mellan provbetyg och slutbetyg. Man kan konstatera att det finns betydande skillnader mellan olika ämnen och mellan olika skolor. I vissa skolor får varannan elev ett betyg som ligger ett steg högre än provbetyget. I andra kan de i genomsnitt få samma eller lägre betyg. Hur skall dessa skillnader tolkas och förstås? Detta är en svår fråga och utredningsgruppen har inget självklart svar.

När det gäller prov som stöd för betygssättning kan man tänka sig två funktioner

- a) proven fungerar som kalibrerande mätinstrument som mer eller mindre strikt styr betygssättningen

Dnr 01-2003:2038

- b) proven fungerar som *stöd* genom att var förebildliga exempel som påvisar hur nationella mål och betygskriterier kan förstås i relation till ett konkret stoff och elevers lösningar.

Punkten a) karaktäriserar traditionella svenska prov av typen standardprov och centrala prov. Det är ett prov för eleverna och elevernas resultat är det centrala.

Denna typ av betygstöd förutsätter, om det skall ingå i ett målrelaterat system, att det finns ett gemensamt väldefinierat kunskapsinnehåll och gemensamma väldefinierade betygsnivåer som både provkonstruktion och undervisning kan utgå ifrån. Ett sådant system har vi inte idag. Som diskuteras i kapitel 4 har vi ett decentraliserat system med en ansvarsfördelning där lärare och elever skall välja innehåll i undervisningen, formulera lokala mål och betygskriterier samt bestämma vilka arbetsmetoder som skall användas för att nå målen.

Det nuvarande systemet bygger alltså på att de lokala aktörerna skall tolka generella och därmed oprecist skrivna nationella mål och kriterier i termer av det undervisningsinnehåll man valt. Det innebär i princip (om än inte i praktiken) att kurserna kan se olika ut i varje enskild skola, med olika mål och olika betygskriterier beroende på stoffval.

Även de grupper som konstruerar de nationella proven gör en tolkning av de nationella målen och kriterierna. Frågan är om denna tolkning och det av provgruppen valda innehållet skall ges högre status än ett lokalt valt innehåll, och om provgruppens tolkning av mål och kriterier skall ses som överordnad en lokal tolkning av mål och kriterier. I det förstnämnda fallet måste svaret vara nej, i det senare är det mer tveksamt, i varje fall om man vill skapa en gemensam norm för betygssättningen.

Man kan konstatera att med den typ av styrdokument Sverige har skulle ett betygskalibrerande provsystem av typ a) vara orättfärdigt och slumpmässigt om skolorna skulle åläggas att följa provresultaten oberoende av hur väl det egenvalda kursinnehållet stämmer med provets innehåll.

En tänkbar lösning som kombinerar en typ a)-modell med nuvarande styrsystem är att proven kopplas till ett tydligt ramverk där det klargörs vad som kan komma att ingå i proven, vilket innehåll som kan tänkas, vilken typ av uppgifter som kommer att användas, hur de kommer att bedömas osv. Men i så fall kommer dessa ”frameworks” att i praktiken fungera som en form av innehållsangivelser och framstå som styrande.

Nationella prov som stöd för betygssättning genom förebildlighet enligt punkt b) är däremot i linje med det nuvarande systemets grundläggande principer. Eftersom proven främst riktar sig till lärarna för att ge förebildliga exempel på hur elevers arbeten kan bedömas behöver proven inte vara obligatoriska. Denna typ av prov läm-

Dnr 01-2003:2038

par sig inte för direkt betygssättning av de egna eleverna eftersom det inte är givet att man valt att arbeta med ett innehåll som ger ett rättvisande utslag på provet. Det är också tveksamt om denna typ av prov lämpar sig för uppföljning. På nationell nivå kan det kanske vara befogat eftersom det ger en översiktlig bild av kunskapsläget i riket. På lokal nivå blir resultaten otillförlitliga som mått på måloppfyllelse om man har arbetat med annat stoff än det som ingår i provet. Däremot kan provresultaten naturligtvis utgöra underlag för konstruktiva diskussioner (förutsatt att proven inte är sekretessbelagda).

Vilken av de två modeller som skissats ovan som gäller eller uppfattas gälla för det nuvarande systemet är oklart.

Utredningsgruppen bedömer att provens betygsstödjande funktion är oklar och svårförståelig. Detta torde också vara fallet bland lärare, elever och föräldrar. Det saknas tydliga instruktioner om hur provresultaten skall förstås och provsystemets roll i skolans ansvars- och styrsystem är oklar och svårbegriplig. Ur detta perspektiv är provsystemet av låg kvalitet.

Provsystemet ska konkretisera kursmål och betygskriterier

Utredningsgruppen ställer sig kritiska till att provsystemet skall ha detta syfte. Kursmål och betygskriterier bör konkretiseras i god tid innan prov skall genomföras. Det är inte rimligt att man skall invänta ett antal nationella prov innan man kan få någon uppfattning om vad staten avser med sina formuleringar av kursmål och betygskriterier. Sekretessen medför vidare att det endast är de lärare som deltar i proven som får ta del av denna konkretisering. Man kan dessutom fråga sig hur de ämnen som inte har nationella prov skall få tillgång till någon konkretisering av sina kursmål och betygskriterier.

Utredningsgruppen bedömer att en konkretisering av kursplaner och betygskriterier är nödvändig för att ge stöd för en likvärdig och rättvis betygssättning. En sådan konkretisering är dock inte viktig bara i de tre provämnena utan i alla ämnen - meritvärden baseras inte bara på provämnena - och bör därför utformas som särskilda kommentarmaterial.

Utredningsgruppen föreslår att kommentarmaterial som bygger på autentiska elevarbeten och utförliga bedömningsanvisningar av det slag som ingår i nationella prov, skapas i alla ämnen

Provsystemets diagnostiska funktion

Provsystemets diagnostiska funktion är förhållandevis oproblematisht och behöver inte närmare kommenteras. De diagnostiska materialen är uppskattade och fungerar bra (även om inte alla känner till dem). Ett problem om man önskar använda de nationella proven i diagnostiskt syfte är att de ges i slutet av utbildningen och det, i synnerhet i gymnasieskolan, inte finns mycket tid kvar för att åtgärda konstaterade

Dnr 01-2003:2038

rade svagheter. Å andra sidan är det svårt att ge ett prov som skall pröva måluppfyllelsen långt före den tidpunkt målen skall vara nådda. Det är alltså svårt att i ett instrument förena ett betygsstödjande och ett diagnostiskt syfte.

Provsystemet ska ge underlag för uppföljning

Utredningsgruppen tycker att det är tveksamt om det nuvarande provsystemet lämpar sig för uppföljning. Som tidigare nämnts ger resultaten möjligen en hygglig nationell bild av en generell måluppfyllelse. De nuvarande proven är dock inte konstruerade för att resultat skall aggregeras och t.ex. jämföras mellan olika år. Stoffvalets betydelse gör också att de inte lämpar sig särskilt väl för jämförelser mellan klasser eller skolor.

Utredningsgruppen ser det som en mycket tveksam investering att göra totalinsamling av resultaten på de nuvarande nationella proven. Insamlingen är sårbar, kostsam och arbetskrävande för skolorna och resultaten är svårtolkade som visats i kapitel 3 i utredningen. Vem skall använda dessa resultat och vad skall de användas till? Om de skall användas för lokala jämförelser anser utredningsgruppen att det är tillräckligt med ett nationellt stickprov till vilket olika skolor kan relatera sina egna resultat inom ramen för det lokala kvalitetsarbetet.

Utredningsgruppen anser inte att värdet av totalinsamling motiverar kostnaden och ställer sig negativ till totalinsamling av provresultat. Utredningsgruppen förordar stickprovsbaserade insamlingar.

Proven styr innehåll och arbetsmetoder

Detta ser utredningsgruppen som en effekt av proven och inte som ett syfte, eftersom någon avsikt att proven skall styra inte finns uttalad (däremot sägs att de skall vara ett *stöd* för betygssättningen). Att proven i praktiken i större eller mindre grad har en styrande effekt ser utredningsgruppen som givet. Det är dock svårt att avgöra hur stark denna styrning är.

Den intressanta frågan är om man skall se denna styrande effekt som önskvärd eller som i strid med det rådande systemet och därmed inte önskvärd. Utredningsgruppen anser att det finns skäl att tydliggöra Skolverkets syn i denna fråga.

I det rådande styrsystemet har staten avsagt sig styrningen när det gäller innehåll och arbetsmetoder. Frågan är vad som då styr i stället. En rimlig tolkning med visst empiriskt stöd från Skolverkets granskningar är att läroböckerna har en klart styrande effekt. Lärarutbildningen har rimligen viss betydelse, och antagligen sker mycken styrning genom trädning, dvs. lärare upprepar det innehåll och de arbetssätt de själva mött under sin skolgång. Det senare torde i särskilt stor utsträckning vara fallet för den ökande andelen obehöriga lärare, för vilka också läroboken kan antas vara kraftigt styrande.

Dnr 01-2003:2038

De nuvarande nationella proven har en sorts retroaktivt styrande roll. När ett antal prov genomförts visar sig efterhand en bild av vad staten anser vara ett lämpligt undervisningsinnehåll och ett lämpligt sätt att arbeta med och bedöma detta innehåll. Detta är ett otydligt och olikvärdigt sätt att styra eftersom alla lärare, elever och föräldrar inte ges samma förutsättningar att ta del av proven (inte minst genom sekretessen).

Styrkan i provens styrande effekt torde vara beroende av vilken typ av betygsstödjande roll de nationella proven tilldelas. Om det är ett kalibrerande betygsstöd med stark koppling mellan provresultat och betyg blir den styrande effekten stark. Om provsystemets indirekt betygsstödjande roll betonas blir styrningen antagligen mindre.

Utredningsgruppen konstaterar att de nationella proven har styrande effekter men tar inte ställning till om detta skall tydliggöras och uttålas som önskvärt.

Proven används som examensprov

Risken för att proven uppfattas som examensprov ser utredningsgruppen som ett förhållande som måste beaktas. Enligt de enkäter som provinstitutionerna ger till provanvändarna och de enkäter som ingått i Skolverkets effektstudie, anser lärarna att de inte i särskilt hög grad styrs av provresultaten i sin betygssättning. Det finns dock ett tryck från olika håll att proven skall få en större betydelse för likvärdigheten och då avses oftast en större följsamhet, dvs. att proven skall få en tydligare kalibrerande funktion.

Vissa resultat i kapitel 3 i utredningen talar också för att provresultaten ges en särskild vikt som de rimligen inte borde ges om de ses som ett av flera likvärdiga betygssunderlag.

Det nuvarande styrsystemet är dessutom komplext och svårförståeligt och flertalet vuxna har själva utbildats i ett skolsystem med tydligt betygskalibrerande prov (även om standardprov och centrala prov var kalibrerande på gruppnivå). Flera studier som Skolverkets kvalitetsgranskningar och Stockholm stads inspektörsrapporter tyder på att de nationella provens roll är oklar bland lärare och elever.

Vad, varför och varthän?

Trots att lärare och rektorer, och till stor del också elever, anser att de nationella proven fungerar bra anser utredningsgruppen att det finns problem och spänningar i det nuvarande provsystemet. Det gäller kanske framför allt frågan om betygens likvärdighet och därmed provens betygsstödjande funktion, men också frågan om provens användbarhet för snabb, tillförlitlig och effektiv uppföljning och utvärdering av skolsystemet på både nationell och lokal nivå.

Dnr 01-2003:2038

I bedömningen av systemets nuvarande funktion har man att ta hänsyn till i vilken utsträckning styrsystemets grundläggande principer skall beaktas. Men man måste också ta hänsyn till behov som inte kan uppfyllas utan att man bryter mot väsentliga delar i det nuvarande styrsystemet, t.ex. i form av inskränkning av det lokala friutrymmet och den deltagande målstyrningen. Detta skapar spänningar i systemet, vilket diskuteras i kapitel 4 i utredningen. Man kan argumentera för att det finns mer djupgående spänningar i skolans styrsystem som får följd effekter på bedömning och betygssättning, t.ex. att styrdokumentet ställer orealistiska krav på professionens förmåga att utan statligt stöd i form av tydliga kommentarmaterial och exemplifieringar, skapa likvärdighet i betygssättningen.

Utredningsgruppen vill för tydlighetens skull peka på två ytterlighetsalternativ till nytt eller förändrat provsystem. Det ena alternativet har ambitionen att så väl som möjligt följa det nuvarande styrsystemets intentioner. Det andra alternativet representerar en pragmatisk och instrumentell ambition att styra mot en likvärdig betygssättning.

Alternativen kan ses som ultimata konsekvenser av de slutsatser och bedömningar utredningsgruppen har gjort av det nuvarande systemets teoretiska utgångspunkter, de förväntningar som läggs på det och dess praktiska tillämpningar.

Alternativ 1: Ett nationellt provsystem med tonvikt på läroplanens intentioner

Premissen för det här förslaget är att valet av stoff och arbetsmetoder skall göras lokalt av lärare och elever. Alternativet kan indelas i tre delar.

1. Kommentarmaterial om bedömning och betygssättning med autentiska elevarbeten som underlag utarbetas i alla ämnen som betygssätts i grundskolan.²

Kommentar: Detta är egentligen inte en del av provsystemet utan snarare en del av styrsystemet. Man kan således mer se förslaget som en ersättning för de nuvarande nationella proven som förväntas ha en liknande roll, dock i retroaktiv mening och endast i de ämnen som har nationella prov. Utredningsgruppen anser att sådana kommentarmaterial är en nödvändig - men måhända inte tillräcklig - förutsättning för att likvärdighet genom förebildlighet skall vara möjlig.

2. Nationella prov av nuvarande slag konstrueras, t.ex. vart tredje år.

Kommentar: Dessa nationella prov kan ses som en form av uppdateringar av kommentarmaterialen. Att proven endast ges vissa år minskar risken för att de skall betraktas som examensprov. Elevernas

² Detta är dock knappast möjligt i gymnasieskolan utan där måste en prioritering göras.

Dnr 01-2003:2038

provresultat behöver inte betygssättas utan det viktiga är att bedömningsmaterialet ger tydliga exempel på hur olika elevarbeten bör bedömas. Proven har alltså en normerande funktion i vid mening. Proven skulle inte vara sekretessbelagda utan tvärtom är det viktigt att de diskuteras så offentligt som möjligt.³

3. Uppföljning genomförs genom att nationella stickprov deltar i utvärderingar inom olika skolämnen. Dessa utvärderingar gäller fler ämnen än de nuvarande provämnena och genomförs enligt ett rullande schema med vissa års intervall för respektive ämne. År 1 kan t.ex. engelska, historia, fysik och bild ingå, år 2 svenska, samhällskunskap, kemi och musik, osv. År 5 återkommer engelska, historia... osv.

Kommentar: Ett sådant uppföljnings- och utvärderingssystem skulle använda samma metodiska uppläggning som Pisa, Pirls m.fl., dvs. med roterade häften så att ett stort kunskapsområde kan provas utan att belastningen på de enskilda eleverna behöver bli särskilt stor. Instrumenten skulle vara sekretessbelagda för att kunna användas upprepat över tid för trendmätningar. Det kan dock användas lokalt av de skolor och huvudmän som så önskar. Man kan också tänka sig att inför inspektioner ålägga skolor att genomföra delar av kunskapsprovningarna för att få ett lokalt underlag som kan bedömas i relation till de nationella resultaten.

Denna typ av instrument och de resultat som erhålls behöver inte uttryckas i betygstermer utan det är möjligt att på samma sätt som för Pisa och Pirls skapa ett system för kvalitetsangivelser, t.ex. enligt ”bookmarkingmetoden”⁴.

Ett uppföljningssystem av den här föreslagna typen skulle kunna uppfylla kraven på lokal och nationell uppföljning. Genom att instrumenten inte är betygsgrundande uppfattas de inte som examensprov och det finns därför inte heller några problem med sårbarhet av det slag som gäller för betygsgrundande nationella prov. Detta är ett provsystem enligt läroplanstyp 3.

Nackdelen med alternativet är att det inte finns något objektivt sätt att med hjälp av instrument bedöma likvärdigheten i betygssättningen. Alternativet bygger på tilltro till att de förebildliga exemplen och de professionella samtalen skall ge tillräckligt underlag för likvärdiga betyg.

³ Dessa prov är egentligen inte nödvändiga i ett system enligt alternativ 1 eftersom man lika gärna kunde tänka sig en uppdatering av kommentarmaterialet vart tredje år. Det kan dock finnas skäl att bibehålla viss provkonstruktion, inte minst för att bevara kompetensen inom provområdet.

⁴ Se bilaga i utredningsunderlaget för närmare beskrivning av det tekniska tillvägagångssättet.

Dnr 01-2003:2038

Alternativet torde möta opposition från lärare, rektorer och kanske elever och föräldrar. Och kanske inte minst från de opinionsbildare som generellt förespråkar ökad användning av prov i olika former.

Alternativ 2: Ett nationellt provsystem för instrumentell kalibrering av betyg och generering av uppföljningsdata.

Detta är ett alternativ som på ett mer handfast sätt än alternativ 1 styr betygssättningen. Det ger också data som kan användas för uppföljning på lokal och nationell nivå. Alternativet kan sägas bestå av två delar.

1. De nationella kursplanerna och betygskriterierna kompletteras med kommentarmaterial där tydliga exempel ges på vad som kan komma att ingå i de nationella proven.

Kommentar: Eftersom proven skall vara styrande för betygssättningen är det av rättviseskäl nödvändigt att lärare och elever har en rimligt klar bild av vad proven kan innehålla. Om inte alla får samma möjligheter att bedöma vad proven innehåller blir slumpinflytandet för stort och en rättvis betygssättning med proven som riktmärke kan inte åstadkommas. Proven måste styras av ett tydligt ramverk.

2. Provdeltagandet är obligatoriskt och provbetyg sätts för varje elev. Betygsnivå och betygsfördelning i klassen får inte avvika alltför mycket från motsvarande resultat på det nationella provet. Prov måste ges varje gång betyg skall sättas.

Kommentar: Det här är i princip samma modell som tidigare användes för standardprov och centrala prov. Skillnaden gentemot dessa är att betygssättningen fortfarande skulle baseras på mål och betygskriterier. Innehållet i proven utgår från ramverkets exemplifieringar av de nationella styrdokumentens generella målformuleringar. Om man talar i de läroplanstermer som diskuteras i kapitel fyra kan man säga att detta är ett provsystem baserat på standards i enlighet med läroplanstyp 4.

De resultat som erhålls i detta system kan användas för lokal och nationell uppföljning eftersom provresultaten är aggregerade på ett standardiserat sätt. En nackdel med systemet är att det får klar karaktär av examensprov, eller i varje fall av s.k. "high stake"-prov. Detta gör systemet sårbart för publicering. Eftersom likvärdigheten vilar på provresultaten behöver prov ges varje år eller termin (i gymnasieskolan). Detta blir ett dyrt system. Av kostnadsskäl torde det då vara nödvändigt att sekretessbelägga proven med påföljd att de inte kan ingå i en pedagogisk - didaktisk diskussion. De kan inte heller visas för elever, föräldrar, lärarutbildare och andra intressenter. Detta måste ses som nackdelar.

Den stora fördelen skulle vara att betygssättningen skulle bli mer likvärdig och enhetlig. Men vi skulle också få ett skolsystem som skulle bli mer likformigt och centralstyrt. Och om denna likformighet

Dnr 01-2003:2038

verkligen ger en mer rättvis betygssättning kan, som diskussionen i kapitel 6 visar, ifrågasätts.

Det här alternativet till provsystem skulle troligen möta motstånd från pedagogiska forskare och från lärare och lärarrepresentanter som anser att proven skulle bli alltför styrande och därmed inskränka den professionella friheten..

Överväganden

Dagens nationella prov befinner sig någonstans mellan de två mer renodlade alternativ som redovisats ovan och måste enligt utredningsgruppen så göra. Med en negativ syn på nuvarande provsystem kunde man säga att det varken representerar det ena eller det andra alternativet. Med en mer välvillig tolkning kunde man kanske säga att det representerar båda alternativen. Och kanske har tiden ännu varit för kort för att det nya betygssystemet och de nationella proven skall ha hunnit få genomslag i det praktiska arbetet.

Det finns behov som är av likartad karaktär oberoende av vilket system man förordar. Det gäller t.ex. behovet av kommenterande och exemplifierande material, antingen beträffande bedömning och betygssättning överhuvudtaget som i det första alternativet, eller beträffande innehåll, utformning och principer för konstruktionen av de nationella proven som i det andra alternativet. I båda fallen handlar det om behov av större tydlighet.

Den stickprovsbaserade uppföljningsmodell som skissas i det första alternativet är givetvis möjligt även för det andra alternativet. Behovet kan dock bedömas som mindre i det andra alternativet beträffande provämnena. I det andra alternativet kommer dessutom provkostnaderna att bli betydligt högre eftersom proven ges varje år eller varje termin och därmed blir det ekonomiska utrymmet för uppföljning i andra ämnen än provämnena mindre med det alternativet.

Utredningsgruppen bedömer att alternativ 1 är det renodlat systemenliga alternativet. Om man skall positionera det nuvarande nationella proven mellan alternativen så ligger det betydligt närmare alternativ 1 än alternativ 2. Det är också rätt påtagligt att de problem man kan identifiera främst rör provens roll för en likvärdig betygssättning. Kapitel 3 i utredningen och andra studier tyder på det även om belägg inte är ovedersägliga utan mer kan ses som indikatorer.

Utredningsgruppen ser allvarligt på den oro som finns om likvärdigheten i betygssättningen, liksom på de behov som finns om konsekvent uppföljning av resultat på lokal och nationell nivå. Det senare behovet är dessutom särskilt betonat i regeringsuppdraget. Med tanke på betygens stora betydelse för behörighet och urval till högre utbildningar är det nödvändigt att betona frågan om provsystemets roll för en rättvis och likvärdig betygssättning. Enligt Skolverkets be-

Dnr 01-2003:2038

dömning måste därför provsystemet utformas så att det på ett bättre sätt än dagens stödjer en sådan likvärdig och rättvis betygssättning.

Samtidigt är det dock väsentligt att provsystemet inte står i alltför uppenbar strid med det nuvarande styrsystemets grundläggande principer. Där betonas de professionellas betydelse och vikten att lärare och elever själva kan välja det ämnesstoff och de arbetsmetoder som på bästa sätt leder till att målen för undervisningen uppfylls. I grunden syftar dessa principer till att skapa så goda lärandemiljöer som möjligt för eleverna.

Utredningsgruppens överväganden om dels behovet av likvärdighet i betygssättningen, dels behovet av frihet och flexibilitet vid val av ämnesinnehåll och arbetsmetoder, resulterar i det kompromissförslag till nytt provsystem som presenteras i nästa avsnitt.

Utkast till nytt provsystem

Vart och ett av de två ytterlighetsalternativ till förändrat provsystem som presenteras i föregående avsnitt bedöms i olika avseende ha så betydande nackdelar att inget av dem kan förordas i renodlad form. Utredningsgruppen ser därför någon form av kompromiss som det enda realistiska alternativet. Följande är ett utkast till ett sådant kompromissförslag till nytt nationellt provsystem. Det behöver givetvis diskuteras och analyseras i sina detaljer och förslaget skall ses som en grov skiss.

Grundskolan och motsvarande skolformer

- Diagnostiska material tillhandahålls i nuvarande omfattning
- Nya nationella prov i år 5 tillhandahålls vartannat eller vart tredje år.

Kommentar: Proven i år 5 är inte betygsgrundande och det finns därför inget nämnvärt behov av sekretess. Samma prov kan användas flera år i följd. Detta förbilligar verksamheten och ger bra underlag för jämförelser över tid.

- Nationella prov tillhandahålls i år 9 i nuvarande omfattning. Proven skall frisläppas efter terminsslut så att innehåll och resultat kan diskuteras offentligt. Proven är obligatoriska men obligatoriet bör kombineras med någon form av rekommendation, instruktion eller föreskrift om hur resultaten på provet skall användas.
- Ett sekretessbelagt reservprov skall finnas. Detta prov skall finnas inlåst hos varje skola och kan användas om skolan bedömer att det ordinarie provet blivit känt. Reservprovet skall inte publiceras utan beräknas vara i funktion t.ex. fem år.

Dnr 01-2003:2038

- Ingen utvidgning till nya ämnen eller åldrar görs för betygsstödjande nationella prov.
- Stickprovinsamling av provresultat.

Kommentar: Ett stickprovsbaserat underlag är tillräckligt för att ge en nationella bild. Denna bild kan sedan tjäna som jämförelse för den lokala bedömningen.

- Ett rullande stickprovsbaserat system för kunskapsutvärdering utvecklas i ämnen som inte har nationella prov. Exempelvis kan tre ämnen utvärderas år 1, tre andra år 2, ytterligare tre andra år tre osv. Efter t.ex. 5 -6 år upprepas cykeln.

Kommentar: I jämförelse med ett system där alla ämnen utvärderas samtidigt t.ex. vart tionde år har den föreslagna modellen fördelarna att utvärderarkompetensen kan upprätthållas och att resultat kan presenteras snabbare. Nackdelen kan vara att alla ämnen inte kan ställas i relation till varandra för samma elevgrupp. Kommuner som själva önskar delta kan använda de nationella instrumenten och jämföra sina resultat med de nationella. Förslaget beskrivs närmare i kapitel 5.

Gymnasieskolan och motsvarande skolformer

- Prov i kärnämnen på samma sätt som ämnesprov i år 9, dvs. årliga (eller terminsvisa) obligatoriska (eller frivilliga) ej sekretessbelagda prov. Reservprov skall finnas om reguljära provet sprids i förtid.

Kommentar: En konsekvensanalys måste göras med avseende på hur frekvent proven skall erbjudas, om de skall vara obligatoriska eller frivilliga, hur lång sekretessen skall vara etc. Gymnasieskolans många olika sätt att organisera undervisningen medför problem när det gäller provens placering i tiden. Även komvux behov av att använda proven behöver utredas närmare. Det är dock viktigt även för gymnasieskolans del att proven inte ingärdas av sekretessregler utan att de kan utgöra underlag för en levande pedagogisk- didaktiska debatt.

- Prov i karaktärsämnen läggs i provbank. De blir därmed sekretessbelagda, men t.ex. vart tredje år kan prov eller provuppgifter och resultat analyseras, publiceras och kommenteras. Om deltagande skall vara obligatoriskt eller frivilligt bör utredas liksom frågan om eventuella krav på resultatredovisningar.

Kommentar: Provbanken erbjuder ett flexibelt, snabbföränderligt och robust system. Om ett prov blir känt kan den enskilda skolan gå in och hämta kompletterande uppgifter i provbanken.

Dnr 01-2003:2038

- Ett rullande stickprovsbaserat utvärderingssystem av samma typ som i grundskolan används för de kärnämnen som inte har nationella prov och för vissa karaktärsämneskurser utöver de som har provbanksprov.

Gemensamt förslag för alla skolformer

Likvärdigheten i betygssättningen måste enligt utredningsgruppens bedömning vila på tre ben a) tydliga utgångspunkter för betygssättningen, b) tydliga provresultat samt c) tydlig uppföljande redovisning av hur provresultaten används. Detta leder till följande förslag.

- Ämnes- eller kursinriktade kommentarmaterial om bedömning och betygssättning bör framställas för flertalet ämnen.

Kommentar: Provsystemet kan enligt utredningsgruppens bedömning inte ensamt bära ansvaret att konkretisera kursplaner och betygskriterier. Flertalet ämnen har därtill inte några nationella prov som kan fylla en sådan uppgift. Förslaget ligger inte inom ramen för det nationella provsystemet men läggs fram därför att det bedöms vara en nödvändig förutsättning för en likvärdig och rättvis betygssättning.

Om ett sådant kommentarmaterial finns kan det utgöra gemensam bas för provkonstruktion och lärarbedömning. En likvärdig betygssättning är därtill viktig i alla ämnen, inte enbart i provämnena. Dessa kommentarmaterial skall vara konkreta och baseras på autentiska elevarbeten. Det skall dock betonas att dessa material skall ses som förebildliga *exempel* och inte som föreskrifter om innehåll och arbetsmetoder. Kommentarmaterialen skall alltså exemplifiera bredden i ämnet eller kursen, inte ses som föreskrifter om likformighet i innehåll eller arbetsmetoder.

- Utredningsgruppen anser att delprovsbetygen skall samordnas till ett provbetyg på ett centralt fastställt sätt.
- Utredningsgruppen bedömer att det inte är möjligt eller önskvärt i rådande skolsystem att på central nivå bestämma att en viss relation skall råda mellan provbetyg och slutbetyg.

Kommentar: Utredningsgruppen bedömer att provbetyg skall finnas och att de skall finnas regler för hur de fastställs. Detta är nödvändigt om provresultat skall vara jämförbara (jämför diskussionen i kapitel 6). Nuvarande ordning för proven i svenska och engelska där varje lärare själv bestämmer hur de olika delprovsresultaten skall användas innebär en godtycklighet som försvårar jämförbarhet.

Däremot förordar utredningsgruppen *inte* att det centralt skall föreskrivas att provbetyg och slutbetyg skall förhålla sig på ett visst givet sätt till varandra. Detta ställer krav på likformiga förhållanden som står i alltför uppenbar strid mot styrdokumentens grundläggande utgångspunkter för att kunna förordas.

Dnr 01-2003:2038

- Utredningsgruppen föreslår att det skall finnas en föreskrift om att skolan och kommunen i sin kvalitetsredovisning skall kommentera sin användning av provresultat och sin syn på relationen mellan provbetyg och slutbetyg.

Kommentar: Detta förslag ligger inte heller inom ramen för provprogrammet men ses som ett nödvändigt tredje ben för att uppnå likvärdighet i betygssättningen. Genom förslaget åläggs skolan och kommunen att aktivt reflektera över likvärdigheten i den egna betygssättningen. Förslaget ger också mening åt det obligatoriska provdeltagandet genom att provresultaten får en koppling till en uppföljande åtgärd. Kvalitetsredovisningarnas beskrivningar av den lokala hanteringen av provresultaten kan sedan diskuteras vid Skolverkets inspektioner. (Jämför punkt B i det aktuella regeringsuppdraget om kvalitetsindikatorer).

Förslagets koppling till det nationella provsystemets syften

Det föreslagna provsystemet

- fyller uppgiften att tillhandahålla diagnostiska material på samma sätt som nuvarande provsystem,
- bidrar till att förtydliga och offentliggöra den nationella tolkningen av kursplaner och betygskriterier genom att proven görs offentliga (Detta medför att stödet till lärarnas eget bedömningsarbete ökar eftersom alla kan ta del av prov och bedömningsmaterial. Offentliggörandet av proven stärker dessutom rätts säkerheten genom att elever och föräldrar kan granska prov och provresultat.),
- stärker likvärdigheten genom att kommuner och skolor åläggs att i sina kvalitetsredovisningar diskutera likvärdigheten i den egna betygssättningen och relationen mellan provbetyg och slutbetyg,
- behovet av nationell och lokal uppföljning uppfylls genom den rullande utvärderingen.

Kostnad

Kostnaderna för förslaget till förändrat provsystem behöver utredas närmare, men en grovskattning kan göras. Om man räknar in nuvarande kostnader för de nationella proven, provbanken och de nationella utvärderingarna (NU 92 och NU 03) så bedömer utredningsgruppen att förslaget till förändrat provsystem är kostnadsneutralt.

De nationella proven för grundskolan och i gymnasieskolans kärnämnen ändras inte. Karaktärsämneshänsyn läggs i provbank vilket leder till lägre kostnader. Systemet med rullande utvärdering och uppföljning av ämnen som inte har prov bedöms rymmas inom den budget som anslagits för nationella kunskapsutvärderingar.

Dnr 01-2003:2038

Utredningsgruppens förslag

Det ingår inte i regeringsuppdraget att ge förslag till nytt nationellt provsystem. Utredningsgruppen ser det dock som viktigt att de olika delarna i regeringsuppdraget om provsystemet bedöms ur ett helhetsperspektiv och föreslår därför att Skolverket framför behovet av ett modifierat provsystem som bättre än dagens kan uppfylla olika syften. Utredningsgruppen föreslår därför att Skolverket ställer sig bakom det kompromissförslag som presenteras ovan.

Provbanken

Distributionen av prov är en kostsam och sårbar del i provhanteringen. För framtiden bör det nuvarande tillvägagångssättet att skicka pappersprov ersättas av elektronisk distribution. Den nuvarande provbanken fungerar som en distributionskanal för vissa prov. För andra prov tjänstgör den mer som ett underlag för utveckling av ”framtidens” prov i virtuella miljöer. I det första fallet handlar det om prov i mer traditionella provämnen som språk och naturvetenskap medan det senare gäller främst prov för yrkesinriktade program. Det handlar i det fallet inte om att tillhandahålla prov som genererar provdata av traditionellt slag utan mer om prov som ger underlag för att bedöma elevers kompetensprofiler utifrån respektive program mål. I vilken utsträckning sådana kompetensprofiler är lämpliga för nationell och lokal uppföljning är i dagsläget svårt att bedöma. Däremot är de värdefulla för det pedagogiska arbetet och som diagnostiska material.

Utredningsgruppen bedömning är att provbanksverksamheten har stor utvecklingspotential. För närvarande är det dock osäkert vad skolornas datautrustning tillåter i form av provdistribution eller för, vilket är nästa steg, onlineprov. De senare innebär att eleverna genomför provet framför en dator. De uppenbara fördelarna med detta är att inga papper behöver skickas, provet kan utnyttja ljud och rörliga bilder, data kan registreras direkt och uppföljningen blir omedelbar. Det är dessutom enkelt att tillhandahålla olika varianter av samma prov vilket minskar sårbarheten och ökar bedömningssäkerheten.

För närvarande kostar provbanksprojektet knappt 15 miljoner kronor. Det är idag svåröverskådligt och omfattande men utredningsgruppen bedömer att provbanker via Internet har stor utvecklingspotential, inte bara för prov utan även för insamling av t.ex. enkätdata, något som redan genomförts inom ramen för NU-03-projektet. Både provverksamhet och enkätinsamling kan också samordnas inom samma system.

I dag ligger de olika provbanksverksamheterna utspridda på olika universitetsinstitutioner. En genomlysning och utvärdering av nuva-

Dnr 01-2003:2038

rande provbanksverksamhet pågår sedan en tid tillbaka. Resultaten är inte klara och några mer konkreta rekommendationer kan inte ges i dagsläget. Det framgår dock att det bör finnas möjligheter till rationaliseringar och samordningsvinster.

Utredningsgruppen rekommenderar att Skolverket uttalar sig för en fortsatt kraftfull satsning på att utveckla provbanksverksamheten.
--

Dnr 01-2003:2038

INLEDNING

Den här rapporten är underlag för Skolverkets redovisning av den del av uppdraget avseende resultatinformation (U2003/2060/S) som gäller det nationella provsystemet, del E. Del E av uppdraget lyder i sin helhet

Skolverket redovisa verkets bedömning av dagens system med nationella prov med avseende på kvalitet och kostnadseffektivitet. Verket bör analysera förutsättningarna för att bredda systemet med nationella ämnesprov till dels fler åldersgrupper, dels fler ämnen eller ämnesgrupper. Utgångspunkt bör vara både den lokala nivån och den nationella nivån behov av uppföljning. Detta bör ske för såväl grundskolan och motsvarande skolformer som för gymnasieskolan och motsvarande skolformer. (sid. 8)

Uppdrag E må synas tämligen enkelt, men det har långtgående implikationer och med den knappa tidsram som stått till buds kan vissa av frågeställningarna inte behandlas mer än översiktligt. Uppdraget innehåller flera komplicerade begrepp, till exempel är det inte självklart vad som ska avses med ett provsystems *kvalitet*. Provsystemet ingår i ett större skolsystem med vilket det växelverkar i olika avseenden. Skolans styrdokument inverkar på provsystemets utformning och tillämpning samtidigt som provsystemet medför tolkningar av styrdokumentet vilket får återverkningar på arbetet i skolan.

Inte heller *kostnadseffektiviteten* låter sig beskrivas eller analyseras på något enkelt sätt. Det är välkänt inom den internationella forskningen om pedagogiska mätningar att konstruktionen av prov av hög kvalitet är en komplicerad, tidskrävande och kostsam procedur. Något som inte alltid är uppenbart för en lekman som kan tycka att prov kan väl varenda lärare konstruera. Provkonstruktion är emellertid ett eget vetenskapsområde med allt vad det innebär av begrepps- och teorbildningar. Det är inte möjligt att i en studie av det här slaget göra några mer ingående beskrivningar och analyser av vad som kan avses med ”kvalitet” i provsammanhang.

Till provsystemets kvalitet kan också acceptansen och legitimiteten hos användarna räknas. Skolverket genomför under 2003 en effektstudie i grundskolan av det nationella provsystemet där synpunkter från lärare, elever och rektorer samlas in. En preliminär redovisning av dessa resultat ingår i rapporten.

Ett viktigt syfte med de nationella proven är att ge underlag för en likvärdig bedömning och en rättvis betygssättning. Vad

Dnr 01-2003:2038

detta kan innebära och hur provresultat och betyg förhåller sig till varandra redovisas och diskuteras i ett eget avsnitt.

Frågan om att *breda* de nationella proven till att omfatta flera ämnen och ämnesgrupper och flera åldrar behandlas i rapporten

I uppdraget anges slutligen att såväl den lokala som den nationella nivåns behov av *uppföljning* ska vara vägledande. Hur detta ska förstås är inte alldeles uppenbart. Provsystemet har och har haft flera mer eller mindre tydligt formulerade syften, varav möjligen uppföljning kan ses som ett. Det har dock inte varit tydligt uttalat i hittillsvarande provuppdrag. En kort redovisning av tidigare och nuvarande provuppdrag inleder därför redovisningen. Förutsättningarna för att använda provsystemet som underlag för uppföljning på olika nivåer kommer att diskuteras i olika sammanhang i redovisningen.

Rapporten har följande disposition. De tre första kapitlen är beskrivningar av det nuvarande provsystemet ur olika aspekter. Kapitlen fyra och fem är mer inriktade mot ett eventuellt framtida provsystem. Det avslutande kapitlet kan ses som ett försök till syntes.

I kapitel 1 ges en kortfattad beskrivning av de hittillsvarande provuppdrag Skolverket haft från regeringen. Vidare ges en beskrivning av dagens provsystem, dess omfattning och var de konstrueras. Ett försök till jämförelser av kostnaden för provutveckling i olika länder redovisas. Slutligen redovisas och diskuteras en del frågor som är kopplade till provens sårbarhet och konsekvenser av de nuvarande sekretessreglerna.

Kapitel 2 redovisar preliminära resultat från en utvärdering om det nationella provsystemet som Skolverket genomfört i grundskolan våren 2003.

I kapitel 3 görs en jämförande analys av provbetyg och slutbetyg. Resultaten diskuteras ur olika aspekter.

I kapitel 4 görs ett försök att kategorisera och finna vissa karakteristiska egenskaper i olika läroplaner. Det sena 1900-talets svenska läroplaner diskuteras med särskilt fokus på frågor som rör kunskapsbedömning och betygssättning.

Kapitel 5 redovisar några erfarenheter och resultat med avseende på en breddning av provsystemet.

I kapitel 6 görs ett försök till sammanfattande analys och diskussion.

Varje kapitel avslutas med en ruta där utredningsgruppen lämnar sin bedömning och i vissa fall ger förslag till åtgärder.

Dnr 01-2003:2038

NUVARANDE PROVSYSTEM

1. Utformning, omfattning och kostnader

Här ges en kort beskrivning av det nuvarande nationella provsystemet – på vilka grunder innehållet byggdes från början, vilken förändring delar i provsystemet har genomgått och det nuvarande systemets omfattning.

Tidigare provuppdrag

Utformningen av det nationella provsystemet enligt regeringens uppdrag 1994-04-21⁵ till Skolverket måste ses mot övriga förändringar i skolsystemet vid den tiden. Övergången från en regelstyrd skola till en i huvudsak målstyrd skola och den ändrade ansvarsfördelningen mellan stat och kommun, när det gäller genomförandet av skolverksamheten, var två genomgripande förändringar. En annan gällde definitionen av kunskapsbegreppet och i förarbetena till de nya läroplanerna talas om det vidgade kunskapsbegreppet. Mycket kort kan här sägas att det handlar om att se kunskap som kontextuell, komplex, mångfasetterad och att se kunskaper som redskap för att lösa problem. En sådan kunskapsdefinition ställer naturligtvis alldeles särskilda krav på det instrument varmed kunskapen ska prövas.

I bilaga till regeringsbeslut⁶ sägs att

Skolverket har av regeringen fått i uppdrag att utarbeta betygskriterier i anslutning till de nya kursplanerna. Härigenom kan en jämförbarhet uppnås i betygssättningen. Särskilt när ett nytt system införs, kan dock principen om en likvärdig betygssättning behöva stödjas med hjälp av centralt utformade prov. ...Det är även viktigt att med sådana fortsätta att ge en så god grund som möjligt för en rättvis betygssättning över landet, eftersom betygen har en stor roll som urvalsinstrument vid intagning till högre utbildning.

Utöver vad som nämnts ovan skulle proven således vara stödjande inför betygssättningen - inte styrande som de hade varit i det tidigare systemet.

Uppdrag 1994⁷

Grundskolan

Vid en jämförelse mellan det tidigare nationella provsystemet med standardprov i grundskolan och centrala prov i den gymnasiala utbildningen och det system som nu skulle byggas upp, var skillnaderna störst i grundskolan. Från att det hade funnits prov i svenska, engelska och matematik vid ett tillfälle i slutet av grundskolan – engels-

⁵ Regeringsbeslut, 1994-04-21, dnr U94/1031/Gru

⁶ Bilaga till regeringsbeslut 1994-04-21, dnr U94/1031/Gru

⁷ a.a.

Dnr 01-2003:2038

ka årskurs 8 och matematik och svenska i årskurs 9, skulle Skolverket nu tillhandahålla diagnostiskt material för åren 2 och 7 samt ämnesprov för skolår 5 och skolår 9. De ämnen det gällde var likt tidigare svenska, engelska och matematik. Ämnesproven för skolår 9 var redan från början obligatoriska för lärarna att använda. Övrigt material ställdes till skolornas förfogande.

Syftet med de diagnostiska proven var att ge lärarna stöd i att diagnostisera elevernas svårigheter som underlag för nödvändiga insatser. Proven under det andra skolåret skulle handla om grundläggande färdigheter i att läsa, skriva och räkna, medan proven under det sjunde skolåret skulle ge lärarna stöd vid bedömningen av var eleverna stod i förhållande till slutmålet.

Ämnesprovets syfte var att kontrollera om eleven uppnått kursplanernas kunskapskrav. Proven för det nionde skolåret skulle också ge underlag för lärarnas betygssättning av elevernas kunskaper.

Allt material för grundskolan skulle även kunna användas i tillämpliga delar i specialskolan, sameskolan och den grundläggande vuxenutbildningen.

Med hänsyn tagen till bl.a. förutsättningarna ovan kom prov och övrigt material för grundskolan att utvecklas enligt följande riktlinjer.

- Strävan var att åstadkomma en kontinuitet genom hela grundskolan, främst med tanke på eleverna. Därför bestämdes att såväl ämnesproven som de diagnostiska proven skulle vila på samma grunder vad gällde förhållningssätt mot skolan/lärarna och vad gällde definitionen av kunskap och tolkning av bedömningsgrunder.
- Skolverket menade att det var lärarna som "ägde" de material som Skolverket ställde till förfogande. Detta rimmade väl med den nya ansvarsfördelning. Det uteslöt inte att lärarna hade vissa skyldigheter att återrapporera resultat och synpunkter.
- Sättet att se på kunskap ledde till nya och mer omfattande elevuppgifter som i sin tur i vissa fall ledde till mer tid för rättningsarbetet än vad man var van vid. Som hjälp tog Skolverket fram s.k. bedömningsunderlag vars innehåll byggde på utfallet i utprovningen. Avsikten var att lärarna skulle få exempel på hur en bedömning av olika kunskaper kunde gå till, och att de sedan själva fick bedöma kvaliteten i sina egna elevers prestationer. Meningen var också att bidra med en terminologi för att mot uppställda mål diskutera och bedöma kvalitativa skillnader i elevernas prestationer.
- Provresultat skulle samlas in från ett urval av skolor med syftet att ta fram en generell rapport för landet som helhet. Tanken var att varje skola skulle kunna diskutera sina resultat i förhållande till de aggregerade resultaten på riksnivå.

Dnr 01-2003:2038

Den gymnasiala utbildningen

Den gymnasiala utbildningen skulle en gång per termin ha tillgång till nationella prov i svenska kurs B, engelska kurserna A och B och matematik kurserna A och E. Utöver kurserna A och E i matematik skulle varje termin finnas ett prov från någon av kurserna B, C eller D.

Vidare skulle Skolverket överväga och redovisa om det var möjligt att erbjuda centralt utarbetade kursprov i B-språk och/eller i naturvetenskapliga ämnen för att stödja en rättvis betygssättning.

Ett skäl till att erbjuda prov två gånger under läsåret var att möta och på något sätt matcha den s.k. fria kursutläggningen i gymnasieskolan.

Proven var frivilliga för lärarna att använda och två datum bildade en genomförandeperiod under vilken proven kunde genomföras. Det första datumet för respektive prov var det datum provet var klart och kunde distribueras till skolorna. Det andra var sista datum för inskick av provresultat enligt uppgjord ordning.

Syftet med proven framgick av gymnasieförordningen (SFS 1992:394) 7 kap.4 § tredje stycket: *Lärarna bör använda centralt fastställda prov som ett hjälpmedel för att bedömningsgrunderna skall bli så enhetliga som möjligt över landet.*

Följande punkter avses lyfta fram några utgångspunkter för konstruktionen av de nationella kursproven för den gymnasiala utbildningen:

- Kursproven skulle vara en naturlig fortsättning på grundskolans prov och övriga material vad gäller innehåll och form. Kongruens och progression utgjorde nyckelbegrepp. Utgångspunkten var således att proven ställdes till lärarnas förfogande, men att de skulle genomföras på det sätt det var tänkt för att ge det underlag som avsågs. Det vidgade kunskapsbegreppet ledde till s.k. breddningsdelar i proven. Dessa innehöll mer omfattande uppgifter som skulle lösas skriftligt eller i vissa fall muntligt.
- I informationen om kursproven klargjordes att proven skulle användas i slutet av en kurs eller efter genomförd kurs. Önskemål från lärarna om att använda proven som diagnoser i början av eller under en kurs avvisades.
- Resultat på proven samlades in från ett urval av skolor med samma syfte som för grundskolan.

Provbanks och nationella utvärderingar

I uppdraget från regeringen⁸ om att utveckla ett nytt nationellt provsystem ingick även att *bygga upp en provbank, ur vilken skolor kunde hämta prov för användning vid den tidpunkt de fann lämpligast.* Uppbyggnaden skulle starta på försök inom några ämnesområden.

⁸ Bilaga till regeringsbeslut 1994-04-21, dnr U94/1031/Gru

Dnr 01-2003:2038

Parallellt med utvecklingen av diagnostiska prov och ämnesprov för grundskolan samt nationella kursprov för den gymnasiala utbildningen påbörjades alltså uppbyggnaden av provbanken. Målet var att allt material skulle lagras och hämtas elektroniskt. Prov i tyska och franska samt i matematik och fysik var de första inslagen i provbanken. Efterhand väcktes frågan om att utveckla prov för program med yrkesämnena. Arbetet startade och ett annorlunda material – jämfört med gängse prov – började ta form.

Den nationella utvärderingen av grundskolan 1992 (NU 92) genererade många och intressanta resultat. I uppdraget till Skolverket⁹ uppmärksammades detta på följande sätt: *Motsvarande nationella kunskapsbedömningar bör även genomföras för gymnasieskolan och den kommunala vuxenutbildningen.*

Uppdrag 1999¹⁰

Bakgrund

I *Utvecklingsplan för förskola, skola, och vuxenutbildning*¹¹ står inledningsvis att mål- och resultatstyrningen ligger fast och ska tillämpas fullt ut. Det torde även innebära att ansvarsfördelningen mellan stat och kommun/huvudman ligger fast och att den utgångspunkten för utveckling av det nationella provsystem fortfarande gäller.

I utvecklingsplanen står vidare att

Det decentraliserade och målstyrda styrsystemet har många fördelar. Det förutsätter dock en god kontroll av resultaten, särskilt de kvalitativa.

Prov eller andra former av kunskapsbedömningar, framtagna på nationell nivå, uppfattas i ett reformskede som en hjälp att tolka kursplanerna och vilka krav de ställer. Den kunskapssyn som läroplanerna anger och som uttrycks på olika sätt i kursplanerna och i betygskriterierna ger helt andra förutsättningar för utvärdering av kunskaper. Provuuppgifter kan inte lägre vara ”enkla” mått av traditionellt slag. De måste också ge möjlighet att analysera vad de som fått en viss utbildning kan göra snarare än att endast redovisa minneskunskaper. De bör ha en inriktning mot problemlösning, tillämpningar och kombinationer av olika kunskapsområden. Inlärningsresultat visar sig mer som övergripande kompetenser och attityder än som faktaredovisningar. Det gäller i hög grad läroplanernas mål och värdegrund men också ämnesmålen i kursplanerna. Detta kräver nya sätt att ta fram underlag och analysera inlärningsresultat.

I propositionen¹² *Gymnasieskola i utveckling, kvalitet och likvärdighet* behandlas de nationella kursproven i den gymnasiala utbildningen. Rubriken är Prov och kvalitetssäkring. Liksom i utvecklingspla-

⁹ Bilaga till regeringsbeslut 1994-04-21, dnr U94/1031/Gru

¹⁰ Regeringsbeslut 1999-09-23, U1999/3290/S.

¹¹ Regeringens skrivelse 1996/97:112.

¹² Regeringens propositionen 1997/98, Gymnasieskola i utveckling, kvalitet och likvärdighet.

Dnr 01-2003:2038

nen¹³ är kvalitet och kvalitetssäkring nyckelbegrepp. För att nå nationell likvärdighet menar regeringen att man i ett decentraliserat system måste ha en systematisk kvalitetssäkring som ska redovisas i särskilda kvalitetsredovisningar, som i sin tur ska ligga till grund för utveckling och förbättring. Det är i denna anda regeringen föreslår att de nationella kursproven i den gymnasiala utbildningen ska bli obligatoriska att genomföra. Det gäller proven i svenska/svenska som andraspråk kurs B, engelska kurs A och matematik kurs A. Därutöver ska obligatoriet gälla den högsta gemensamma karaktärsämneskursen i engelska och matematik på respektive program. Det innebär engelska kurs A för dem som läser denna kurs och matematik kurserna B, C och D för dem som läser respektive kurs som högsta gemensamma kurs.

Ett annat kvalitetssäkringsresonemang i nämnda proposition är att alla som ska genomföra ett prov ska ha samma möjligheter att förbereda sig – egentligen att inte förbereda sig genom att se provet i förväg. Regeringens slutsats är, att ett visst nationellt kursprov i princip är användbart vid endast ett enda tillfälle. Det innebär också att alla i hela riket ska genomföra provet vid ett och samma tillfälle.

Grundskolan

För grundskolans del innebär det nya provuppdraget att de diagnostiska proven ändrades. Det första var att namnet ändrades från diagnostiska prov till diagnostiska material. I och med att materialen enligt det första provuppdraget var destinerade till skolår 2 och skolår 7, hade somliga den uppfattningen att de gestaltade uppnåendemål för dessa skolår – uppnåendemål som inte fanns då heller. Uppgiften för Skolverket blev att ”sträcka ut” diagnosmaterialen så att de kunde användas dels för åren före skolår 6, dels för åren 6 till 9. Syftet med materialen skulle vara att ge lärarna hjälp att bedöma och stödja elevernas utveckling. Materialen för åren före skolår 6 skulle i första hand ha fokus på språkutveckling och matematisk begreppsbyggnad. Syftet med materialen för skolår 6 t.o.m. 9 var att de bör ge underlag för bedömning av var varje elev står i förhållande till uppställda mål.

Åren före skolår 6 inkluderar även förskolan och förskoleklassen. För dem bl.a. utvecklades ett s.k. analyschema i matematik, vilket i korthet innebär att pedagogerna kan se ett schema över hur elevens matematikkunnskap utvecklas under skolåren, vilka delar som ingår och vilka steg som tas. Med den kunskapen kan de sedan se när olika barn tar de första stegen i sin matematikutbildning och gå in och förstärka eller lägga tillräkta.

En motsvarighet till analyschemat i matematik finns i svenska och heter då observationsschema.

¹³ Regeringens skrivelse 1996/97:112.

Dnr 01-2003:2038

Den enda förändringen beträffande ämnesproven för skolår 5 och 9 var att svenska och nytilkomna svenska som andraspråk skulle provas med samma prov, men med kompletterande kommentarer i bedömningsunderlaget.

Den gymnasiala utbildningen

Regeringens förslag att göra de nationella kursproven obligatoriska för lärarna att använda antogs och 7 kap. 4 § tredje stycket gymnasieförordningen¹⁴ fick följande lydelse:

Inför betygssättningen i kärnämnen svenska, svenska som andraspråk, engelska och matematik skall lärarna använda sig av nationellt fastställda prov. Nationellt fastställda prov skall också användas i den avslutande gemensamma karaktärsämneskursen i engelska och matematik. Även i övrigt bör lärarna använda nationellt fastställda prov som ett hjälpmedel för att bedömningsgrunderna skall bli så enhetliga som möjligt över landet.

Den förändring som blev nödvändig var att i matematik flytta kurs E-provet till provbanken och att ta fram prov för kurserna B, C och D varje termin. Vidare beslutade Skolverket om provdatum, vilka kungjordes i SKOLFS. Här upplevdes ingen större förändring ute i skolorna, eftersom tidigare SKOLFS också hade ett datum, dvs. det datum proven var klara, distribuerade och kunde användas. Genom SKOLFS gavs rektor möjligheten att besluta om ett senare provdatum i den händelse provet ej kunde genomföras utsatt datum.

Den kommunala vuxenutbildningen berördes inte av gymnasieskolans obligatorium. Där är proven fortfarande frivilliga och *lärarna bör använda nationellt fastställda prov för att bedömningsgrunderna skall bli så enhetliga som möjligt över landet*¹⁵.

Provbanken

I uppdraget angavs följande riktlinjer för vidareutvecklingen av provbanken.

- Provbanken skall i först hand omfatta ämnen och kurser i vilka det inte finns nationella prov.
- Provbanken skall innehålla prov eller delar av prov samt diagnostiskt material som kan utgöra ett stöd för läraren vid bedömning av elevernas prestationer.
- Provbankens tekniska och innehållsmässiga uppbyggnad skall syfta till en flexibel användning för samtliga berörda skolformer. Skolor skall kunna utnyttja provbanken vid för dem passande tillfällen.

¹⁴ Gymnasieförordning, 1992:394

¹⁵ Förordning om kommunal vuxenutbildning, 1992:403, 4 kap. 6 §, tredje stycket.

Dnr 01-2003:2038

- Provbanken bör utformas så, att den erbjuder olika forum, t.ex. genom Internet, för diskussion mellan lärare om provbankens utformning och användning.

Angående vad som idag ingår i Skolverkets provbank hänvisas till översikten i nästa avsnitt.

Sammanfattande kommentar

Den grund på vilken proven byggdes från början har inte förändrats under de år det ”nya nationella provsystemet” varit i bruk. I regeringens utvecklingsplaner som regeringen har skrivit har betonats att mål- och resultatstyrningen ligger fast, det vidgade kunskapsbegreppet har utvecklats och ansvarsfördelningen mellan stat och kommun/huvudman har inte ändrats. Det som har tillkommit är den alltmer omfattande diskussionen om kvalitet och kvalitetssäkring och skolhuvudmannens ansvar att bedöma kvaliteten i den egna verksamheten för att därefter rapportera till Skolverket. Det har då legat nära till hands att se elevernas resultat på nationellt fastställda prov som ett mått på kvalitet och den fråga som rests i sammanhanget är i vilken grad resultaten är tillförlitliga. Tveksamheten har bl.a. bottnat i att den enskilde lärarens omdöme har varit en del av resultatet, trots att detta var en del av själva poängen med proven från början.

Man skulle kunna säga att det finns en motsättning idag mellan den pedagogiska, stödjande och utvecklande ambitionen med det nationella provsystemet och den kontrollerande och styrande ansatsen med krav på tillförlitliga data som går att aggregera på olika nivåer. Det är tveksamt om ett nationellt provsystem kan rymma båda dessa syften. Om bedömningen är att så inte är möjligt, torde slutsatsen bli att en ambition måste vika.

Provverksamheten läsåret 2002/2003

Syftet med det här avsnittet är att ge en sammanfattande bild av det nuvarande provsystemets omfattning.

Grundskolan

Grundskolans ämnesprov tillhandahålls vårterminen under aktuellt läsår. För läsåret 2002/2003 innebär det att prov för år 9 tillhandahålls enligt följande tabell.

Dnr 01-2003:2038

Tabell 1. Prov för år 9 våren 2003

Ämne	Delprov	Tid	Provdatum		
Svenska och Svenska som andraspråk	A (läs)	80 min	5	februari	2003
	B (munt)	valfri		valfri tid	2003
	C (skriv)	160 min	7	februari	2003
Engelska	A (munt)	15-25 min		valfri tid	2003
	B (recept)	135 min	13	mars	2003
	C (skriv)	80 min	21	mars	2003
Matematik	A (munt)	ca 25 min		valfri tid	2003
	B (uppg)	130 min	9	april	2003
	C (tema uppg)	80 min	29	april	2003

Varje prov består av tre delprov (A, B och C) varav ett är muntligt och kan genomföras vid valfri tidpunkt. Övriga delprov genomförs vid i tabellen angivna tidpunkter. Även provtiden för de olika delproven anges.

För år 5 tillhandahålls prov i samma ämnen som i år 9. Dessa prov är frivilliga och något provdatum är inte fastställt utan proven kan användas fortlöpande under vårterminen.

Proven i engelska och matematik år 9 är generellt sekretessbelagda i tio år. Proven i svenska däremot är endast sekretessbelagda den termin de ges. Syftet med den tioåriga sekretessen är tvåfaldigt, dels är avsikten att sänka kostnaderna för nykonstruktion av uppgifter, dels är det värdefullt att kunna använda samma uppgifter vid olika tidpunkter för att få en bild av hur elevgruppens kunskaper förändras över tid. Proven distribueras i tryckt form till varje skola i de antal skolorna beställt.

I matematik anges poänggränser för ett provbetyg. I svenska och engelska däremot anges fr.o.m. läsåret 2001/02 endast underlag för betyg på de olika delproven. Någon metod eller princip för att erhålla ett samlat provbetyg anges således inte. Lärarna får själva tolka det samlade provresultatet.¹⁶

Proven i år 5 är för närvarande avsedda att kunna användas i två år. De nykonstrueras således vartannat år. Provet utgörs av en pärm med kopieringsunderlag och distribueras i ett exemplar till varje skola utan beställning.

Till detta kommer att diagnostiska material tillhandahålls i svenska och matematik för åren upp till årskurs 6 och i dessa två ämne plus

¹⁶ Beslut togs efter en intern diskussion av GD 2001-06-20. Motiveringen var att det är lärarens uppgift att sätta betyg och att staten inte ska styra den processen.

Dnr 01-2003:2038

engelska för årskurserna 6 till 9. Även dessa material distribueras utan beställning i ett exemplar till varje skola.

I samtliga fall är det Liber distribution som sköter tryckning och distribution.

Grundskolans nuvarande provverksamhet omfattar således

År 9	3 obligatoriska prov per år, nykonstruerade men med inslag av återanvända uppgifter.
År 6-9	3 diagnostiska material som förnyas när så bedöms lämpligt.
År 5	3 frivilliga prov som byts ut vartannat år.
År F-5	2 diagnostiska material som förnyas när så bedöms lämpligt.

Gymnasieskolan

Provverksamheten i gymnasieskolan är betydligt mer omfattande. Det har sin grund i att proven är obligatoriska i kärnämneskurserna i samma tre ämnen där det finns prov i grundskolan, samt att de är obligatoriska i de avslutande karaktärsämneskurserna i samma ämnen på respektive program. Ytterligare skäl är att prov tillhandahålls varje termin samt att det tillhandahålls provbanksprov i ett ytterligare ett antal ämnen utöver de tre tidigare nämnda. Anledningen till att prov ges vid flera tillfällen under läsåret är att provsystemet inte ska inkräkta på skolornas möjligheter att förlägga kurser till olika terminer.

För läsåret 2002/2003 tillhandahålls prov i gymnasieskolan i enlighet med tabell 2.

Dnr 01-2003:2038

Tabell 2. Nationella kursprov för gymnasieskolan läsåret 2002/2003.

Ämne	Provdatum			Provtid
Engelska kurs A (muntl)		valfri ht	2002	15 min/par
(höra, läsa, skriva)	18	november	2002	210 min
Engelska kurs A (muntl)		valfri vt	2003	15 min/par
(höra, läsa, skriva)	6	Maj	2003	210 min
Engelska kurs B (muntl)		valfri ht	2002	15 min/par
(höra, läsa, skriva)	22	november	2002	210 min
Engelska kurs B (muntl)		valfri vt	2003	15 min/par
(höra, läsa, skriva)	15	maj	2003	210 min
Matematik kurs A	29	november	2002	180 min
Matematik kurs A	16	maj	2003	180 min
Matematik kurs B	12	december	2002	240 min
Matematik kurs B	21	maj	2003	240 min
Matematik kurs C	10	december	2002	240 min
Matematik kurs C	23	maj	2003	240 min
Matematik kurs D	6	december	2002	240 min
Matematik kurs D	19	maj	2003	240 min
Svenska kurs B / Svenska som andraspråk kurs B (munt)		valfri ht		5 min/elev
(skriv)		valfri ht		80 min
(skriv)	12	november	2002	300 min
Svenska kurs B / Svenska som andraspråk kurs B (munt)		valfri vt		5 min/elev
(skriv)		valfri vt		80 min
(skriv)	11	april	2003	300 min

Som framgår av tabell 2 tillhandahålls i gymnasieskolan 14 nationella kursprov under ett läsår. Proven för gymnasieskolan är sekretessbelagda på samma sätt som grundskolans prov, dvs. i tio år för proven i engelska och matematik, och under aktuell termin för svenska.

Förutom de ovan angivna proven, som är obligatoriska enligt de föreskrifter som tidigare nämnts, erbjuds ett antal provbanksbaserade prov. Tabell 3 visar de aktuella proven läsåret 2002/2003.

Dnr 01-2003:2038

Tabell 3. Prov från provbanken läsåret 2002/2003

Ämne	Första provdatum	
Franska kurs B/Franska steg 4	20 november	2002
Franska kurs B/Franska steg 4	28 april	2003
Tyska kurs B/Tyska steg 4	20 november	2002
Tyska kurs B/Tyska steg 4	28 april	2003
Matematik kurs E	4 december	2002
Matematik kurs E	14 maj	2003
Fysik kurs A	16 december	2002
Fysik kurs A	8 maj	2003
Fysik kurs B	9 december	2002
Fysik kurs B	12 maj	2003
SFI	Tidigare givna prov	

Till de 14 nationella proven i obligatoriska ämnen ska således läggas 10 prov från provbanken (förutom proven för SFI som är tidigare givna prov).

Utöver de i tabell 3 angivna proven pågår utvecklingsarbete med provbanksprov i ytterligare ett antal kurser och program. I det sist-nämnda fallet gäller det material för att bedöma måluppfyllelse inom ett antal yrkesinriktade program. Följande kurser och program ingår i utvecklingsarbetet

- Biologi A
- Spanska, steg 3
- Franska, steg 3
- Tyska, steg 3
- Industriprogrammet (IP)
- Elprogrammet (EC)
- Medieprogrammet (MP)
- Byggprogrammet (BP)
- Energiprogrammet (EN)
- Hotell- och restaurangprogrammet (HR)
- Naturbruksprogrammet (NP)
- Omvårdnadsprogrammet (OP)
- Handels- och administrationsprogrammet (HP)
- Barn- och fritidsprogrammet (BF)
- Hem- och konsumentkunskap (grundskolan)
- Demokratisk kompetens (grundskola, gymnasieskola)

Sammantaget ingår 16 kurser och program i det pågående utvecklingsarbetet. Till de verksamheter som anges ovan ska läggas den

Dnr 01-2003:2038

tekniska utvecklingen av provbanken som handhas i särskild ordning men bedrivs i samarbete med de institutioner som utvecklar proven.

I sammanfattning omfattar gymnasieskolans provverksamhet varje läsår

Nationella kursprov i obligatoriska provämnena	14
Prov som erbjuds från provbank	11
Provbanksprov under utveckling	16

Sammantaget utarbetas alltså årligen 39 prov som vänder sig till gymnasieskolan (om man räknar bort provet i SFI och i Hem- och konsumentkunskap från provbanken).

Vuxenutbildningen och övriga skolformer

Tanken med provsystemet är att de nationella proven för grundskolan och gymnasieskolan, liksom materialet i provbanken, också ska användas inom den kommunala vuxenutbildningen och i tillämpliga delar inom särskolan, sameskolan osv. De enda provmaterial som konstrueras under 2003 med särskilt inriktning mot vuxenutbildning är prov för SFI

2 nykonstruerade prov per år samt

bedömningsmaterial utifrån nya kursplaner 2002.

Vilka konstruerar proven?

Skolverket ansvarar för provverksamheten, men samtliga prov konstrueras på Skolverkets uppdrag vid olika universitetsinstitutioner. Följande institutioner är kontrakterade.

Dnr 01-2003:2038

Tabell 4. Institutioner engagerade för provutveckling

Institution	Prov
Enheten för pedagogiska mätningar, Umeå universitet	Matematik: prov för kurs B – D, Provbank: matematik E, fysik A, B, biologi A
IML., Umeå universitet	Teknisk utveckling av provbank
Institutionen för nordiska språk, Uppsala universitet	Svenska/svenska som andraspråk: diagnostiska material, Äp 5, Äp 9, kurs B
Prim-gruppen, Lärarhögskolan i Stockholm	Matematik: diagnostiska material, Äp 5, Äp 9, kurs A, Provbank: programprov för programmen IP, EC, MP, BP, EN samt prov i Hem- och konsumentkunskap
Enheten för språk och litteratur, Göteborgs universitet	Engelska: diagnostiska material, Äp 5, Äp5, kurs A, B. Provbank: tyska, franska, spanska
Institutionen för pedagogik, Växjö universitet	Provbank: programprov för programmen HR, NP, OP, HP, BF samt prov i demokratisk kompetens.
Centrum för tvåspråkighetsforskning Stockholms universitet	Prov och bedömningsmaterial i svenska för invandrare (SFI)

Den sammanlagda kostnaden för provverksamheten under 2003 är budgeterad till knappt 45 miljoner kronor. I nästa avsnitt redovisas kostnaderna mer i detalj.

Kostnad, sårbarhet och sekretess

I regeringsuppdraget sägs att ”Skolverket bör redovisa verkets bedömning av dagens system med nationella prov avseende kvalitet och kostnadseffektivitet”. Detta uppdrag kan synas enkelt men rymmer många komplikationer, vilket torde framgå när rapportens olika delar ses i ett sammanhang.

Provsystemets kvalitet och kostnaden för provsystemet kan inte ses som skilda från varandra. Även när det gäller provsystem innebär högre kvalitet högre kostnad. Innan kostnadens rimlighet eller kostnadseffektiviteten kan bedömas måste det därför bestämmas vilka kvalitetskriterier som ska gälla för ett prov eller provsystem. Sådana kriterier är i sin tur beroende av vilka syften provsystemet har. Som framgår i andra delar av rapporten är provsystemets syften inte särskilt tydligt formulerade, och i den mån de formulerats kan de framstå som svårförenliga. Detta gör en bedömning av provens kvalitet - inte minst deras tekniska kvalitet -svår, eftersom en sådan bedöm-

Dnr 01-2003:2038

ning är beroende av att syftena är tydliga. I *Standards for educational and psychological testing*¹⁷ anger standard 13.2

In educational settings, when a test is designed or used to serve multiple purposes, evidence of the test's technical quality should be provided for each purpose.

Det är inte möjligt att inom ramen för den här utredningen gå igenom vilka tekniska kvalitetskrav de nationella provens olika mer eller mindre uttalade syften ställer. Istället hänvisas till rapportens appendix där en mer generell översikt ges av begrepp, teoretiska utgångspunkter samt metodiska och instrumentella överväganden kopplade till kunskapsbedömning baserad på prov. Denna begränsning gör att någon mer ingående analys av kostnadseffektiviteten i relation till provsystemets kvalitet inte kan göras. Vad som däremot kan göras är att beskriva vilka moment som ingår i arbetet med att framställa och distribuera proven samt vilka kostnader som är förknippade med detta. För att få något underlag för bedömning av kostnadens rimlighet kommer också vissa jämförelser att göras med motsvarande kostnader i andra länder.

Viktiga underlag för bedömningen av kostnaden för provprogrammet är också vilka hot systemet utsätts för i form av att prov offentliggörs i förväg, samt i vilken utsträckning prov och provuppgifter kan återanvändas. Oberoende av vilka åtgärder som väljs får detta konsekvenser av olika slag.

I det här avsnittet görs ett försök att beskriva och diskutera kostnaden för provutveckling i relation till kvalitetskrav, sårbarhet, behov av att sekretessbelägga prov och vilka konsekvenser olika åtgärder får.

Provs kvalitet

Utvecklingen av prov ska styras av provens syften och av hur provresultaten ska användas.¹⁸ Utvecklingsprocessen innefattar många olika delar, t.ex. bestämning av innehåll i provet, uppgiftsformat, de förhållanden under vilka proven ska genomföras, vilka konsekvenser resultaten får i olika avseenden, hur de kommer att användas och hur de kan användas.

I arbetet med att utveckla ett nationellt prov ingår således att fastställa ett ramverk för provverksamheten där syfte, innehåll, resultat användning etc. fastläggs. Vidare innehåller ramverket specifikationer för konstruktion av uppgifter, för utprövning, för hantering av utprövningsdata, för utformning av bedömningsunderlag. Därtill ska läggas utformning av rutiner för att pröva uppgifternas tekniska kva-

¹⁷ *Standards for educational and psychological testing* är framtagen och fastställd gemensamt av American Educational Research Association, American Psychological Association och National Council on Measurement in Education och kan ses som vetenskapssamhällets kvalitetsnorm för provframställning och provanvändning.

¹⁸ *Standards for educational and psychological testing*, sid. 37.

Dnr 01-2003:2038

liteter, riktlinjer för hur uppgifter ska väljas till ett helt prov, hur man bör se på relationen mellan mål och uppgift, hur bedömning av olika svar i relation till mål och betygskriterier bör ske, hur samlade provresultat ska hanteras, vilka rutiner som ska tillämpas för att bestämma betygsgränser osv.

Att konstruera prov av hög kvalitet är således en komplicerad och utdragen procedur som kräver kreativitet för att konstruera uppgifter, ämnesdidaktisk kompetens för att bedöma uppgifter i relation till kursmål och betygskriterier, teknisk kompetens för att analysera utprovningsdata etc. Arbetet innebär också täta kontakter och nära samarbete med olika grupper av lärare som deltar vid utprovningar av uppgifter eller ingår i bedömningsgrupper av olika slag.

Hur lång tid behövs för att utveckla prov?

Hur lång tid tar det då att utveckla ett prov som uppfyller ovanstående kriterier? James Popham (2002) som är en välkänd forskare inom området anger i egenskap av ordförande för ”The Commission on Instructionally Supportive Assessment” att

A state must generally allow test developers a minimum of three years to produce statewide tests that satisfy the Standards for Educational and Psychological Testing and similar test-quality guidelines. (sid. 9)

GAO¹⁹ (2003) anger att det tar approximativt två till tre år att utveckla ett nytt prov.

För de svenska nationella proven brukar en utvecklingscykel uppskattas till 1,5 till 2 år, vilket kan sägas ligga väl i linje med vad som anses rimligt av välrenommerade bedömare.

Kostnader för provkonstruktion

Som framgått ovan innefattar utvecklingen av ett nationellt prov av hög kvalitet en rad olika moment. Kostnaden för att utveckla ett nationellt prov beror som tidigare nämnts på vilka kvalitetskrav som läggs på arbetet. Eftersom det knappast är aktuellt att argumentera för sänkt kvalitet är det rimligt att utgå från prov som följer de kvalitetskriterier som Standards for Educational and Psychological Testing anger och som i huvudsak också kan anses giltiga för de nuvarande svenska proven.

Kostnaden kan delas upp i olika delar. En gäller själva utvecklandet av proven, en annan tryckning och distribution, en tredje kostnaden lokalt för själva genomförandet, en fjärde gäller kostnaden bedömning av svaren, en femte insamling och publicering av resultaten.

I många länder rättas prov centralt av särskilt anställda rättare. I Sverige genomförs rättning och bedömning av lärarna själva och denna

¹⁹ United States General Accounting Office.

Dnr 01-2003:2038

kostnad belastar således inte själva provsystemet (även om det naturligtvis är en kostnad lokalt).

De jämförelser som görs här innefattar endast kostnader som rör utvecklingen av proven, dvs. fram till att ett tryckfärdigt prov föreligger. De kostnader som är förknippade med tryckning, distribution och redovisning av resultat redovisas separat.

USA

Den pedagogiska forskningen är mest omfattande i USA och det är också lättast att hitta underlag därifrån. Något enhetligt nationellt provsystem finns inte i USA.²⁰ Totalkostnader för provsystem är svåra att jämföra, beroende på att olika kostnader kan ingå i olika länder, så därför är det "utvecklingskostnad per prov" som är mest intressant i det här sammanhanget. Det kan dock vara intressant att nämna att NASBE (National Association of State Boards of Education) uppskattar den årliga kostnaden åren 2003 till 2008 för prov i matematik och "reading" till mellan 385 miljoner och 1 miljard dollar, dvs. till mellan cirka 3 och 8 miljarder kronor (Legislative brief, june 2003). Detta är dock en skattning och NASBE påpekar att det råder brist på forskning om kostnader för provverksamhet.

I den tidigare nämnda studien från United States General Accounting Office (GAO, 2003) redovisas genomsnittlig utvecklingskostnad per prov för sju delstater. Utvecklingskostnaderna varierar kraftigt från 190 000 kr i Maine till 1,62 miljoner i Massachusetts. Medelvärdet för de sju staterna blev cirka 725 000 kr.²¹

Skillnaden mellan stater kan delvis förklaras med vilken typ av prov man använder. Prov som i hög utsträckning baseras på flervalsfrågor (ett vanligt format i USA) är i allmänhet något billigare att utveckla än prov mer öppna frågor där eleverna ska formulera fullständiga svar. Däremot har staternas storlek ingen betydelse för utvecklingskostnaden. Det är lika dyrt att utveckla ett prov för 50 000 elever som för 500 000 (däremot stiger givetvis andra kostnader som tryck distribution etc.).

De ovan angivna utvecklingskostnaderna gäller prov som redan är i drift. Med tanke på att en del av uppdraget också är att utreda förutsättningarna för att vidga provprogrammet till fler ämnen och åldrar kan det vara intressant att notera att GAO-studien uppskattar att kostnaden för utveckling av nya prov är två till tre gånger högre än för befintliga prov.

I en annan studie har Rebarber och McFarland (2002) uppskattat att utvecklingskostnaden per prov för varje årskurs och ämne till \$ 500

²⁰ Undantaget är NAEP, National Program for Educational Progress, som är ett nationellt stickprovsbaserat program för kunskapsutvärdering. Se appendix för mer information.

²¹ Beräknat efter en kronkurs på 8,50 kr per dollar.

Dnr 01-2003:2038

000, dvs. drygt fyra miljoner kronor. Även i det här fallet rör dig sig om utveckling av nya prov.

Både de reella utvecklingskostnaderna för olika stater och de skattningar av kostnader som olika organisationer och forskare gör varierar som synes påtagligt. Den slutsats man kanske i första hand kan dra är att det är svårt att skatta kostnaden för att utveckla ett prov.

Kanada

Inte heller Kanada har något nationellt provsystem. Däremot har man ett nationellt program för utvärdering av kunskaper i "mathematics content and problem-solving, reading and writing, and science" kallat SAIP (The School Achievement Indicators Program). Programmet gäller 13- och 16 åringar. Projektet drivs sedan 1993 av "the Council of Ministers of Education", som är en samorganisation för de olika provinsernas och territoriernas utbildningsansvariga.²²

Kostnaden för programmet uppges vara 3 miljoner kanadensiska dollar, varav hälften finansieras centralt och hälften lokalt. Den lokala delen torde i huvudsak röra ersättningar för rättning av proven varför utvecklingskostnader, resultatredovisning etc. torde utgöra cirka 1,5 miljoner dollar. En omräkning till kronor gör då att utvecklingskostnaden skulle uppgå till cirka 8,5 miljoner kronor. Om detta slås ut på de sex delarna i programmet (tre ämnen för två åldersgrupper) blir kostnaden per prov cirka 1,4 miljoner kronor. I detta ingår då troligen inte bara utvecklingskostnader utan också kostnader för resultat-sammanställning, vilket i t.ex. Sverige görs centralt av Skolverket och inte av de institutioner som konstruerar proven.

Nederländerna

I Nederländerna har man ett nationellt provsystem sedan 1986. Det sköts på utbildningsdepartementets uppdrag av CITO (Netherlands National Institute for Educational Measurement), ett tidigare statligt institut som 1999 privatiserades (CITO, 2001).

Grundskoleproven består av två delar, en generell del som testar "cognitive level" och en del som är mer tydligt ämnesorienterad. Den generella delen prövar läsförståelse på modersmålet, hörförståelse i engelska, matematik och ett antal generella förmågor. Den innehåller således delar som ingår också i de nationella proven i Sverige. Den ämnesorienterade delen prövar kunskaper i franska som främmande språk, geografi, ekonomi osv. Syftet med dessa prov är dock inte att stödja en betygssättning utan att "help students choose a suitable packade of subjects to study in senior secondary". Sammanlagt konstrueras prov i 15 olika ämnen.

Provkonstruktionen följer till stor del samma mönster som den svenska med noggranna utprovningar och analyser av utprovningsre-

²² Informationen hämtad från <http://www.cmec.ca/saip/indexe.stm>

Dnr 01-2003:2038

sultat. I nederländerna liksom i Sverige är det lärarna som rättar elevernas prov utifrån anvisningar som Cito tillhandahåller.

Kostnaden för "Basic secondary education test" uppgavs 2001 vara 4,5 miljoner gulden. I december 2001 motsvarade 1 gulden 4,27 kronor. Kostnaden för de 15 proven skulle då bli 19,2 miljoner kronor. Räknat gör det en kostnad på 1,28 miljoner kronor per prov.

Cito konstruerar också prov för motsvarigheten till gymnasieskolan. Även där i 15 ämnen men till följd av att det finns olika inriktningar finns sammanlagt 200 varianter av de 15 ämnesproven. Kostnaden för denna del av provprogrammet var 25 miljoner gulden eller knappt 107 miljoner kronor. Räknat på 200 prov skulle det innebära ett genomsnitt på drygt 0,5 miljoner kronor per prov. Det är dock svårt att bedöma hur mycket de olika varianterna av prov i samma ämne skiljer sig åt.

Norden

Förfrågningar har riktats till de övriga nordiska länderna för att försöka få en uppfattning vad deras provsystem och prov kostar. Från Danmark, Finland och Norge har uppgifter erhållits. Uppgifterna baseras på personlig kommunikation med personer inblandade i respektive lands provverksamhet.

Danmark

Svaret från Danmark är kortfattat och anger endast att utgiften för gymnasieskolan är cirka 80 000 danska kr till "opgavekommision" och 10 kr per styck i tryckkostnad.

För grundskolans år 9 uppges att kostnaden för uppgiftskonstruktion i skriftlig framställning likaledes är 80 000 kr och att kostnaden för layout, tryckning, upphovsrätt etc. uppgår till 200 000 kr. Den totala kostnaden bedöms vara 300 000 kr för 105 000 tryckta prov.

Kostnaderna för att konstruera prov i Danmark är således en bråkdel av kostnaden USA och Nederländerna. Det danska provsystemet baseras främst på lärares provkonstruktion: Någon centraliserad provutveckling av det slag vi har i Sverige, och som sker i USA och Nederländerna, tycks inte förekomma.

Norge

I Norge är man i färd med att införa ett system med nationella prov. Dessa konstrueras liksom i Sverige vid olika universitetsinstitutioner. Eftersom verksamheten är relativt nystartad är det svårt att ge någon mer exakt kostnad, men den uppskattning som görs att utvecklingen av prov i genomsnitt kostar 1,35 miljoner svenska kronor.

Finland

Dnr 01-2003:2038

Finland har inga nationella prov av det slag Sverige har på grundskolenivå. Man genomför endast stickprovsbaserade kunskapsutvärderingar med syfte att ge en bild av det nationella kunskapsläget.

Kostnaden för utvärderingen i modersmål i åk 9 uppgick 2001 till drygt 150 000 euro. Räknat på en kronkurs av 9,00 kr skulle det motsvara 1,35 miljoner kronor. I den summan ingår tryckning och distribution samt framställning av en rapport, vilket inte inräknats i de svenska kostnaderna. Eftersom proven endast ges till ett stickprov av elever (3000 - 4000) är dock tryck- och distributionskostnaderna inte särskilt stora. Om man räknar bort 250 000 kr för denna kostnad och rapportskrivning blir kostnaden cirka 1,1 miljoner kr.

På gymnasienivå har man ett omfattande examenssystem men för detta har några uppgifter om kostnad inte erhållits.

Sverige

Givetvis varierar kostnaderna för provutveckling mellan olika prov. Samtidigt konstruerar all provinstitutioner flera olika prov och det är svårt att få fram kostnaden för varje enskilt prov. Det närmaste man kan komma utifrån nuvarande redovisningsformer är en skattning av genomsnittskostnaden för ett prov konstruerat vid en viss provinstitution.

Svenska

Proven i svenska och svenska som andraspråk konstrueras vid institutionen för nordiska språk vid universitetet i Uppsala. Uppdraget innefattar prov för år 5, som för närvarande nykonstrueras vartannat år, och år 9 där ett nytt prov konstrueras varje år. I gymnasieskolan erbjuds prov i kärnämnet svenska B. Det sistnämnda erbjuds i skilda versioner både höst- och vårtermin. Sammantaget innebär det att tre och ett halvt nytt prov konstrueras per år. Till detta ska också läggas diagnostiska material som utvecklas successivt.

För 2003 är kostnaden för provutvecklingen i svenska 4 700 000 kr. Om man räknar med fyra prov per år ger det en genomsnittlig kostnad per prov på knappt 1,2 miljoner.

Engelska

Proven i engelska konstrueras vid Institutionen för pedagogik och didaktik vid Göteborgs universitet. Ersättningen för 2003 uppgår till 8 919 450 kr. Denna summa gäller dock inte enbart proven i engelska utan även provbanksprov i franska, tyska och spanska. Av det anslagna beloppet går 6 520 500 kr till proven i engelska.

I engelska konstrueras liksom för svenska prov för år 5 och 9 i grundskolan. För gymnasieskolans del tillhandahålls prov för kurserna engelska A och engelska B. Även dessa erbjuds både vår och hösttermin. Diagnostiska material utvecklas med viss oregelbunden-

Dnr 01-2003:2038

het för skolåren 6 till 9. Sammantaget innebär det att cirka sex prov tillhandahålls per år. De prov som erbjuds i gymnasieskolan under höstterminen utgörs dock till stor del av återanvända uppgifter och provdelar. Om man därför skattar antalet nykonstruerade prov per år till fem skulle den genomsnittliga kostnaden bli cirka 1,3 miljoner kronor.

Matematik

Proven i matematik konstrueras dels av PRIM-gruppen vid Lärarhögskolan i Stockholm och dels vid Enheten för pedagogiska mätningar vid Umeå universitet. Grundskolan prov samt proven för kärnämneskursen A konstrueras i Stockholm medan kursproven för kurserna B till D konstrueras i Umeå.

PRIM-gruppen konstruerar grundskoleprov och diagnostiska material enligt samma mall som gällde för svenska och dessutom två prov för kurs A. Sammantaget gör det 3,5 prov per år plus visst arbete med diagnostiska material.

Ersättningen till PRIM-gruppen för 2003 uppgår till 5 103 000 kr. Om man utgår från fyra nykonstruerade prov ger det en genomsnittskostnad på knappt 1,3 miljoner per prov. Detta är möjligen något lågt räknat eftersom viss återanvändning av uppgifter sker (cirka 20-25 procent av uppgifterna enligt uppgift från PRIM-gruppen).

Vid Umeå universitet konstrueras kursprov som erbjuds både höst och vår för kurserna B – D, sammantaget sex prov per år. För 2003 är ersättningen 4 859 000 kr, vilket ger en genomsnittlig kostnad per prov på lite över 0,8 miljoner. Även i matematik finns möjligheter till återanvändning av uppgifter. Enligt uppgift är det dock endast i undantagsfall som uppgifter återanvänds. Om man skattar att i genomsnitt ett av de sex proven består av återanvända uppgifter och räknar på fem nykonstruerade prov blir genomsnittskostnaden per prov cirka 1,0 miljoner.

För samtliga uppdrag gäller att det ingår visst arbete med lärarenkäter och rapportering av resultat. Detta kan dock ses som en naturlig del i arbetet med att konstruera prov och kan därför räknas in i utvecklingskostnaden.

Svenska för invandrare, Sfi

Centrum för tvåspråkighet vid Stockholms universitet konstruerar årligen två prov i svenska för invandrare. Ersättningen är 1 700 500 kr vilket gör en kostnad på cirka 850 000 kr per prov.

Provbanksprov

Cirka 14 miljoner av den totala kostnaden för provprogrammet går till utveckling av probbanksverksamheten. Denna innefattar, som tidigare redovisats, ett antal olika ämnen och program. Det är för

Dnr 01-2003:2038

provbankens del inte möjligt att inom ramen för det aktuella uppdraget göra någon mer ingående bedömning av kostnadseffektiviteten.

Provbanksystemet är för närvarande under utredning men det är för tidigt att i det här skedet dra några mer vittgående slutsatser. Några grova skattningar av kostnaderna och deras fördelning kan dock göras, men bilden är splittrad och svåröverskådlig eftersom vissa prov är i en utvecklingsfas och har knappast kommit i praktiskt bruk medan andra har funnits under längre tid.

Provbanken i matematik E och fysik A och B som handhas av Umeå universitet har funnits sedan 1997. Senare har biologi A tillkommit. Provbanken tillhandahåller färdiga prov i matematik och fysik med möjligheter att byta ut och ersätta vissa uppgifter. I biologi finns tillgång till vissa uppgifter men något färdigt prov finns inte för närvarande. Sex prov per år tillhandahålls plus uppgifter i biologi. För 2003 är ersättningen 6 057 000 kr vilket grovt räknat gör cirka 1 miljon per prov. Provbanken syftar dock till att bygga upp en bank av uppgifter för framtida bruk varför en kostnad per prov inte är riktigt relevant.

Provbanken i moderna språk som utvecklas av Göteborgs universitet tillhandahåller prov och provdelar fortlöpande under året. Även den har funnits sedan senare delen av 1990-talet. Ersättningen för proven i tyska, franska och spanska (nytt) uppgår 2003 till 2 398 950 kr. En grovskattning att fyra prov kan användas per år (tyska, franska), att utvecklingsarbete pågår i spanska samt att visst diagnostiskt material i engelska tillhandahålls skulle innebära att kostnaden per prov kunde skattas till cirka 0,5 miljoner.

Övriga prov i provbanken kan ses som utvecklingsprojekt och det är inte meningsfullt att där försöka beräkna någon kostnad per prov.

Distribution och tryck

Kostnaden för tryck och distribution av prov är rätt betydande. För närvarande tillhandahålls fullständigt provmaterial kostnadsfritt för skolorna endast för proven i år 9. Proven i år 5 och de diagnostiska materialen distribueras i ett exemplar till varje skola. Dessa kan sedan beställa fler exemplar om så önskas.

I gymnasieskolan tillhandahålls ett exemplar av respektive prov kostnadsfritt för skolan. Denna kan sedan välja att köpa färdigtryckta prov till eleverna eller själve trycka upp dem.

Tryckning och distribution uppgår med nuvarande rutiner till cirka 3,5 miljoner kronor per år. Om av olika skäl, t.ex. för att minska sårbarheten i provsystemet, färdigtryckta prov skulle skickas till alla aktuella gymnasieelever beräknas kostnaden bli tredubblad., dvs. komma att uppgå till cirka 10,5 miljoner kronor per år.

Om dessutom år 5-prov ska tillhandahållas till alla elever stiger kostnaden ytterligare.

Dnr 01-2003:2038

Slutsatser

Man kan konstatera att kostnaden per prov är rätt likartad för de olika ämnena och inte heller skillnaden mellan institutionerna är särskilt påtaglig. Om kostnaden per prov ska anses hög eller låg är svårt att uttala sig om. Jämfört med de exempel som har redovisats för olika länder ter sig kostnaderna som rimliga. Det är rätt slående hur nära varandra utvecklingskostnaderna i olika länder ligger. Detta betyder givetvis inte med nödvändighet att det nuvarande svenska systemet är det mest kostnadseffektivt, men som nämndes inledningsvis skulle en mer ingående utredning av kostnadseffektiviteten kräva insatser som inte är möjliga inom ramen för det aktuella uppdraget.

Kostnaden kan också ses ur ett annat perspektiv än kostnad per prov. Om man utgår från proven i grundskolan som genomförs av cirka 100 000 elever så skulle kostnaden för konstruktionen av prov bli cirka 12-13 kronor per prov (eftersom utvecklingskostnaden ligger mellan 1,2 och 1,3 miljoner). För en grundskoleelev som genomfört samtliga prov i år 5 och år 9 skulle den samlade kostnaden då bli mindre än 100 kr. Även om man lägger på någon tia för de diagnostiska materialen och uppskattar trycknings och distributionskostnaden till ungefär samma belopp som utvecklingskostnaden (vilket är väl tilltaget) så skulle den samlade kostnaden per elev för provprogrammet bli under 300 kr. I jämförelse med att en grundskoleelev kostar cirka 62 500 kronor *per år* framstår denna kostnad som mycket försumbar.

Om det däremot gäller t.ex. prov D i matematik som endast är obligatoriskt för elever på naturvetenskapliga programmet stiger givetvis kostnaden per elev. Läsåret 2001/02 gick drygt 16 000 elever det naturvetenskapliga programmet. Kostnaden för dessa prov uppskattas till cirka 1 miljon. Om man delar detta på 16 000 elever blir kostnaden per elev 63 kronor. I det här fallet betalas tryckning och distribution av den beställande skolan så kostnaden för Skolverket blir i stort sett 63 kronor per elev. Med tanke på att en genomsnittlig gymnasieelev kostar drygt 70 000 kr per år framstår inte heller detta som någon särskilt avskräckande kostnad.

De nationella provens sårbarhet

Att de nationella provens innehåll inte bör vara känt i förväg torde alla instämma i. Det betyder att proven måste hållas hemliga fram till genomförandet. Detta förutsätter i sin tur att proven hanteras på ett säkert sätt av alla som är involverade i hanteringen av dem, alltifrån uppgiftskonstruktörer, universitetsinstitutioner, tryckerier, distributörer, till rektorer och lärare ute på skolorna.

Men provsystemet är sårbart. Den snabba tekniska utvecklingen har förändrat förutsättningarna. Det är idag ett ögonblicksverk att med en modern digitalkamera fotografera ett prov och lägga ut det på Internet, eller att med hjälp av en kamerautrustad mobiltelefon ta en bild

Dnr 01-2003:2038

och skicka den dit man önskar. Likaså är det ett ögonblicksverk att skicka ljudlösa SMS-meddelanden från och till t.ex. en provlokal. Så frågan är hur ett provsystem bör utformas för att bli tillräckligt robust för att bevara trovärdigheten utan att för den skull kostnaderna blir orimliga. Och utan att hanteringen signalerar examensprov.

Grundskolan

Provmaterialet distribueras dels i januari då mer allmän information skickas ut, dels ett par veckor för provets genomförande då själva proven distribueras.

På morgonen den 9 april 2003, samma dag som delprov B i grundskolans ämnesprov i matematik ska genomföras, ringer en lärare och meddelar att provet finns utlagt på Internet. Sådana rykten har florerat tidigare år, men den här gången kunde en adress anges. En snabb kontroll visade att provet låg på den uppgivna adressen i form av en ”jpg-bild”, dvs. ett foto av provet. Denna bild hade lagts ut kvällen före och kunde snabbt spåras till en grundskolelevs hemsida. Skolan kontaktades och efter att rektor talat med eleven framkom att han fått bilden från en adress i en annan del av Sverige. Ägaren till denna angav att han i sin tur hämtat den från en s.k. IP-adress i Stockholm. I det läget bedömdes att det inte var meningsfullt att leta vidare.

Under dagen ringde flera lärare och skolledare och sammanlagt motogs ett 30-tal telefonsamtal och mejl i frågan. Skolverket gick via brev och hemsidan ut med information om vad som inträffat och uppmanade skolorna själva att bedöma hur de skulle agera. Om man bedömde att proven varit kända var det nödvändigt att vara försiktig i tolkningen av resultaten. Från Skolverkets sida bedömdes att det inte fanns tillräckliga skäl för att inhibera provet. Det finns sammanlagt cirka 1500 skolor med elever i år 9 och det var omöjligt att centralt bedöma i vilken utsträckning eleverna känt till att provet varit publicerat och vilken effekt publiceringen haft på resultaten.

Ett annat skäl till att inte agera med indragning eller ogiltigförklaring av provet var att detta skulle ge signal om att provet och provresultatet skulle ha en alldeles särskild betydelse. I läroplanen, liksom i Skolverkets kommentarmaterial (Skolverket, 2001), betonas att det är läraren som ska sätta betyget och att det nationella provet endast är ett underlag bland andra i den bedömningen. Att då ogiltigförklara proven skulle skicka fel signal. Mycket tyder på att de nationella proven tilldelas alltför stor betydelse²³ och en indragning av provet skulle på ett dramatiskt sätt förstärka en sådan syn. Det var inte önskvärt.

Den 29 april skulle delprov C i matematik genomföras. Ganska snart kom signaler om att även detta delprov var publicerat. Den här gången var skolorna förvarnade och ett antal skolor där man misstänkte

²³ Se t.ex. kapitlet om provbetyg och slutbetyg.

Dnr 01-2003:2038

att elever kände till att provet fanns på nätet valde många att använda ett delprov från tidigare år.

Inte heller när det gäller delprov C var det möjligt att få en bild av hur utbredd kännedomen om provet var. För att få en viss uppfattning om hur skolorna reagerat på det inträffade genomfördes en mindre studie i form av intervjuer med rektorer, lärare och elever vid ett tiotal skolor.

Någon större dramatik hade det inträffade inte orsakat på de besökta skolorna. Man hade noterat vad som hänt och fattat sina beslut och därmed ansågs saken utagerad. Lärarna ansåg i allmänhet att de har så god kännedom om elevernas prestationer att resultatet på det nationella provet inte har särskilt stor betydelse. Det bekräftar oftast vad man redan vet.

Vissa upprörda föräldrar hade hört av sig. Eleverna var mer störda och på alla besökta skolor ansåg elever att det var fusk att få tillgång till proven i förväg, i synnerhet om det leder till att de fuskande eleverna får högre slutbetyg än de annars skulle ha fått.

De aktuella skolornas agerande sedan det blivit känt att proven varit publicerade på nätet varierade. Vissa skolor bedömde att spridningen av proven inte berörde deras elever och genomförde proven i vanlig ordning. Några skolor valde däremot att byta ut det publicerade provet mot ett prov från tidigare år.

Även den interna hanteringen av provmaterialet varierade påtagligt mellan skolorna. I någon skola på en liten ort där alla känner alla lämnade brevbäraren personligen provet till expeditionen eller till rektor. Denne överlämnar sedan proven till ”pedagogerna” som därefter ansvarar för säkerheten. Vid detta tillvägagångssätt kan det förekomma att prov förvaras i olåsta utrymmen, eller till och med att de kan ligga framme på ett skrivbord.

Vid ett par andra skolor låses proven in i plåtskåp och lärarna får dem först i samband med provgenomförandet. Proven räknas och vid insamlingen kontrolleras att alla prov återlämnats.

Sammantaget visar undersökningen att hanteringen av provmaterialet varierar betydligt mellan skolor, alltifrån tydliga, strikta rutiner till en tämligen vårdslös hantering.

Gymnasieskolan

I maj 2003 ringde några gymnasieskolor och informerade om att det kommit till skolornas kännedom att prov i matematik fanns till salu. Det enda konkreta exemplet utgjordes dock av att en elev vid en skola blivit ertappad av rektor när han försökte kopiera ett prov på skolans kopiator.

Händelserna fick dock stor uppmärksamhet i media, inte minst genom att TV gjorde reportage som sändes både i regionala och natio-

Dnr 01-2003:2038

nella nyhetsprogram. I det här fallet kunde inga prov återfinnas på nätet och de uppgifter som lämnades gällde endast att prov varit till salu. Några skolor bytte ut det aktuella provet men i huvudsak torde proven ha genomförts enligt de ursprungliga planerna.

Även om några påtagliga skadeverkningar inte kunde konstateras innebär dock uppgifterna att proven finns tillgängliga att provens legitimitet skadas på ett oacceptabelt sätt. Det är framför allt olyckligt om eleverna uppfattar att de inte bedöms och betygssätts på ett rättvist sätt och det är dessutom dyrt att tillhandahålla prov. Skolverket bedömde det som nödvändigt att markera hur allvarligt verket ser på händelser av det slag som förevarit. En polisanmälan inlämnades därför den 19 juni 2003.

Rättsläget

Rättsläget kring proven är dock inte alldeles enkelt. Enligt utslag från kammarrätten kan nationella prov bedömas vara av den karaktären att de faller under sekretesslagen. Denna gäller emellertid inte elever utan enbart personal. Det betyder att endast om det kan påvisas att någon person varit försumlig i sin tjänsteutövning och därigenom brutit mot sekretesslagen kan åtal väckas.

För elever gäller däremot inte sekretesslagen och det är därmed svårt att med tillämpning av något juridiskt regelverk motivera påföljd för en elev som spridit provet. Det är alltså nödvändigt att hanteringen av prov är så säker att proven inte kommer i händerna på elever eller andra personer som inte faller under sekretessbestämmelserna.

Den 14 augusti meddelade polisen att någon polisutredning inte skulle göras eftersom utredningen inte förväntades leda till åtal.

Förslag till åtgärder för att minska sårbarheten

De åtgärder Skolverket vidtagit med anledning av det inträffade är att förtydliga instruktionerna till främst rektor men också till lärarna att proven måste hanteras med vederbörlig försiktighet.

Vad som på sikt kan göras är en svårare fråga att besvara. Som tidigare nämnts har den tekniska utvecklingen gjort spridningsrisken ofantligt mycket större på bara några år. Några åtgärder är tänkbara.

a) Förseglade försändelser som får öppnas först på provdagen Detta är den metod som tidigare användes för standardprov centrala prov och som fortfarande används av t.ex. Högskoleprovet.

Fördelar: Risken för att prov ska falla i orätta händer minskar. Om så skulle ske skulle det upptäckas genom att försändelsen är bruten.

Nackdelar: Själva inplastningsförfarandet gör provhanteringen dyrare. Lärarna får inte möjlighet att se proven och bedöma om anpassning av provsituationen behöver göras för elever med funktionshinder. Förfarandet skickar signaler om att provet är "high stake", dvs.

Dnr 01-2003:2038

av särskild betydelse. Om provet uppfattas som särskilt betydelsefullt ökar också trycket på att försöka komma över dem. För gymnasieskolans del, där idag endast ett exemplar av provet tillhandahålls gratis, skulle förfarandet antingen bli att skolornas åläggs att köpa färdiga prov, eller att Skolverket tillhandahåller prov till alla aktuella elever. Detta senare förfarandet beräknas innebära merkostnader för Skolverket på cirka sju miljoner kronor årligen.

b) Proven konstrueras i några olika varianter. Varje prov föreligger i t.ex. tre olika men likvärdiga versioner.

Fördelar: Det blir inte lika attraktivt att försöka få tillgång till ett prov eftersom man inte vet vilken version som är aktuell för egen del.

Nackdelar: Det blir dyrare att framställa flera versioner av samma prov.

c) Proven distribueras vi Internet. Detta är i princip den modell som gäller för provbanken.

Fördelar: Distributionen blir enkel eftersom en mängd pappersprov och annat material inte behöver skickas via post. Det är enkelt att hantera olika versioner av samma prov. På sikt kan det bli möjligt att genomföra proven direkt vid dator, varvid olika multimedia-applikationer kan användas. Resultathanteringen kan då också automatiseras så att både enskilda resultat och aggregerade resultat föreligger omedelbart efter det att provet genomförts.

Nackdelar: Det är oklart hur användarnamn och lösenord kan hanteras på ett säkert och smidigt sätt. Oklart om skolorna har tillräcklig datorkapacitet och om skolornas system för tryckning och kopiering av material är tillräckligt tillförlitligt.

Det är knappast någon tvekan om att den sistnämnda metoden är den som på sikt är mest attraktiv. Det för dock för närvarande oklart i vilken utsträckning nackdelarna kan hanteras.

Hur länge ska proven vara hemliga?

Att prov måste vara sekretessbelagda före provtillfället behöver knappast diskuteras. Men hur lång tid efter genomförandet bör de vara hemligstämplade? För närvarande är som tidigare nämnts proven i matematik och engelska sekretessbelagda i 10 år, medan proven i svenska och svenska som andraspråk endast är det under den termin de ges. Vilka är skälen för sekretessen och vilka är konsekvenserna?

Dnr 01-2003:2038

Produktkvalitet och processkvalitet

De tidigare studier som gjorts av provverksamheten tyder, liksom här redovisade resultat, på att proven röner stor uppskattning bland lärarna.²⁴ Genomgående anses de vara av hög kvalitet.

Även de norska granskarna Tobiassen och Thomassen som 1999 gjorde en utvärdering av de svenska ämnesproven i år 9 kommer fram till att proven har, med deras uttryck, hög ”produktkvalitet”. Däremot är de mer tveksamma till provens ”processkvalitet”. Med det senare avses det genomslag som proven får i den lokala och nationella pedagogiska och ämnesdidaktiska debatten. Proven anses således fungera bra som instrument för bedömning och betygssättning men mindre bra som bränsle för den professionella utvecklingen inom det ämnesdidaktiska fältet.

En viktig anledning till att det förhåller sig på det sätt som Tobiassen och Thomassen beskriver torde vara de sekretessregler som gäller. För närvarande är som nämnts den generella regeln att prov i matematik och engelska sekretessbeläggs i tio år medan prov i svenska endast är sekretessbelagda till terminens slut. Genom sekretessen är det t.ex. inte möjligt att i en uppföljande rapportering av provresultaten diskutera resultaten utifrån provens innehåll och utformning. Inte heller är det möjligt att i offentliga rapporter diskutera bedömningsanvisningarna och de överväganden som legat till grund för dessa. Endast statistiska fördelningar av betyg kan publiceras med nuvarande regler. Sådan information kan givetvis vara av värde men knappast när det gäller den ämnesdidaktisk utvecklingen.

Varför sekretess?

Skälen till att det införts sekretessregler är flera. För det första finns det en historisk förklaring genom att de nationella proven när de infördes anpassades till den ansvarsfördelning som säger att skolor och kommuner själva avgör hur skolarbetet organiserades. En följd av detta blev att proven inte borde vara fixerade till en dag utan att de skulle kunna genomföras under en längre period för att tillåta olika kursuppläggningar.

Detta upplägg ledde snabbt till protester. Tilltron till proven blev låg genom att det ansågs att elever kunde få tillgång till dem via elever vid andra skolor där proven genomfördes tidigare än vid den egna. En följd av detta blev att det utvecklades en praxis att de flesta valde att göra provet den första dagen under den angivna provperioden. Regelverket anpassades så småningom till denna praxis och en bestämd provdag infördes.

²⁴ Peter Stern (1999). *Effekter av ämnesprov i år nio*. Skolverket (dnr 97:2218), Tobiassen och Thomassen (2000). *Det nationella provsystemet – sett med norska ögon*. Rapport till Skolverket

Dnr 01-2003:2038

För komvux del gäller dock inte detta och för att proven även ska kunna användas där behöver de vara sekretessbelagda även efter provdagen. Beskrivningen ovan gäller främst gymnasieskolan. För grundskolans del har det funnits fastlagda provdagar sedan starten men de nuvarande sekretessreglerna infördes i samband med proven vårterminen 2000.

Att proven behöver kunna användas vid komvux medför dock inte att de behöver sekretessbeläggas i tio år. För gymnasieskolans del tillhandahålls prov varje termin, vilket innebär att en sekretess under aktuell termin är tillräcklig även för komvux. Om grundskolans prov används vid grundläggande vuxenutbildning kan en sekretess som endast gäller aktuell termin innebära att proven är offentliga under höstterminen. För vuxenstuderande som skulle behöva använda grundskoleprov under höstterminen skulle därför en sekretess som endast gäller vårterminen innebära nackdelen att proven knappast kan användas under hösten.

Ett andra skäl till sekretessregeln var att eftersom stora pengar läggs ner på att konstruera prov och provuppgifter bör kostnaden kunna sänkas om uppgifter eller delar av provet kan användas flera gånger, antingen i nya prov eller i en provbank. För att sådana återanvända prov ska få legitimitet ansågs det nödvändigt att uppgifterna och proven sekretessbelades.

Ett tredje skäl var och är att det finns ett intresse av att följa kunskapernas utveckling och förändring över tid. För att göra sådana trendmätningar är det en stor fördel att kunna använda samma uppgifter med vissa års mellanrum.

Konsekvenser av sekretessen

De angivna skälen för att sekretessbelägga proven är givetvis legitima. Nackdelen är dock att den utomordentliga ämnesdidaktiska utvecklingspotential som nationella prov har inte kan utnyttjas annat än marginellt. Processkvaliteten blir låg. När det gäller proven i svenska och svenska som andraspråk, som inte är sekretessbelagda, finns flera belägg för hur värdefullt det är att i kompetensutvecklingssammanhang kunna arbeta med aktuella prov. I dessa ämnen kan också proven och provresultaten diskuteras öppet i fackpressen vilket måste ses som viktiga inslag i lärarnas professionella utveckling.

Mot sekretesskravet kan också anföras att synen på kunskap och vilka kunskaper som är väsentliga ständigt förändras. Tio år gamla uppgifter känns ofta otidsenliga och i synnerhet om nya kursplaner tillkommit under mellantiden torde flertalet prov och uppgifter inte längre uppfattas som optimala eller ens lämpliga.

En kompromisslösning kunde vara att proven t.ex. gjordes i två delar där en del frisläpps medan en del är sekretessbelagd. Det främsta motivet för detta skulle vara att den frisläppta delen kan ingå i den

Dnr 01-2003:2038

efterföljande ”processen” medan den sekretessbelagda delen kan användas för trendmätningar.

Om den nationella provverksamheten i ökande grad läggs in i ett provbankssystem med olika versioner av samma prov, på det sätt som tidigare diskuterats, kan det också finnas skäl att sekretessbelägga vissa delar av proven. De nationella provens roll i den professionella ämnesdidaktiska utvecklingen är dock så viktig att starka skäl talar för att vissa särskilt intressanta delar av proven bör frisläppas och diskuteras i anslutning till att provet genomförs.

Utredningsgruppens bedömning

De nuvarande provsystemet håller genomgående hög kvalitet i tekniskt avseende. Konstruktionsarbetet följer i huvudsak internationell standard på området.

Jämförelser med kostnaden för provproduktion i olika länder tyder på att kostnadsbilden i Sverige är väl i linje med motsvarande kostnader i flertalet andra länder.

Den genomsnittliga kostnaden per grundskoleelev för ett prov är cirka 12 kronor plus tryckkostnad och frakt, sammantaget mindre än 20 kr. En grundskoleelev kostar 62 500 kr per år, i jämförelse är provkostnaden försumbar.

Den nuvarande sekretessen på tio år ses om ett betydande hinder för provens roll som underlag för en levande pedagogisk-didaktisk debatt.

Det nuvarande provsystemet är mycket sårbart för spridning på t.ex. Internet. En skärpning av regelverket och dess tillämpning är nödvändig. Nackdelen med detta är att proven kan uppfattas som ännu viktigare än i dag och därmed ytterligare öka trycket på att komma över dem. Vidare kan de i högre grad än i dag komma att uppfattas som examensprov.

Sårbarheten innebär även betydande nackdelar genom att totalinsamlingen av provresultat tappar i trovärdighet. Kostnaden för denna insamling framstår som tveksam med hänsyn tagen till riskerna och värdet av de insamlade resultaten.

2. Vad anser användarna om provsystemet?

Inledning

I följande kapitel belyses det nationella provsystemet från ett användarperspektiv. Hur proven fungerar i praktiken och i verksamheten kan ses som en väsentlig del i en bedömning och diskussion av ett systems kvalitet. I någon mån kan det också bidra till värderingen av systemets effektivitet. Kapitlet vilar på resultat av en pågående studie av det nationella provsystemets effekter med avgränsning till grundskolan. Av detta följer att redovisningen inte innefattar gymnasieskolan och att de resultat som redovisas är preliminära.

Som nämns i rapportens inledande kapitel är det inte självklart vad som ska avses med ett provsystems kvalitet. Begreppet kvalitet i sig självt är ett begrepp som kan, bör, och också har problematiserats åtskilligt. I detta avsnitt definieras kvalitet som hur det nationella provsystemet används och fungerar, uppfattas och värderas av olika grupper av användare. I denna definition ligger också frågan om provens legitimitet, så som den uppfattas av användarna. I kapitlet redovisas resultat från en enkät som besvarats av rektorer på ett representativt urval av landets grundskolor och på en enkät som har besvarats av lärare på samma skolor som undervisar i nationella provämnen för år 5 och år 9²⁵. Den baseras också på, men i mindre utsträckning, de intervjuer som har genomförts på ett mindre urval om fyra 1-9 skolor med rektorer och lärare (år 5 och år 9) och på genomförda gruppsamtal med elever år 9 i anslutning till provtillfällen.

En utgångspunkt för presentationen av användarperspektivet är att söka spegla åtminstone del av den variation som det empiriska materialet innehåller. Förutom att presentera studiens huvudsakliga resultat, vad majoriteten lärare, rektorer, elever säger om de nationella provens värde och funktion, syftar avsnittet därför att lyfta fram de olika erfarenheter och uppfattningar som existerar bland olika grupper lärare, rektorer m fl. Som exempel på några undergrupper har skillnader mellan lärare som undervisar i år 5 eller i år 9 genomgående beaktats liksom mellan lärare som svarar för de nationella proven

²⁵ Totalt ingick ett representativt urval av 740 grundskolor varav 379 skolor med ansvar för år 5 och 361 skolor med ansvar för år 9. Till respektive skolas rektor ställdes en enkät och bifogades ett skattat antal enkäter att distribueras till undervisande lärare vårterminen 2003 (enligt vissa angivna principer i följebrevet) i de nationella provämnena, d v s matematik, svenska, svenska som andraspråk, och engelska. Varje lärare svarade enbart för ett ämne i enkätens frågor (även om han eller hon undervisade i fler, t ex i svenska och engelska). Svar har inkommit från 603 rektorer och från 2499 lärare, vilket representerar ca 82 % av skolorvalet och grovt skattat ca 75 % av lärarurvalet.

Dnr 01-2003:2038

i olika ämnen. En jämförelse har också gjorts mellan mer och mindre erfarna lärare.

Om det nuvarande provsystemets kvalitet – erfarenhet och värdering från skolfältet

De aspekter om den nationella provverksamheten som inleder kapitlet avser provens möte med skolan, d v s det praktiska hanterandet med proven ute på skolorna och vad det får för konsekvenser exempelvis för lärares tid och för undervisningens planering. Därefter följer en beskrivning av hur viktiga proven ses vara för verksamheten och hur väl provens olika syften ses som uppfyllda av lärare och rektorer. Provens eventuella påverkan, effekt på undervisningens innehåll och form och mer generellt på skolans verksamhet fokuseras och ett ytterligare avsnitt ställer frågan hur man på skolnivå använder sig av de nationella provresultaten. Slutligen fokuseras elevmaterialet där de samtal som förts med grupper av elever i år 9 läggs samman med den indirekta bild som framkommer om prov-elev-relationen i lärares svar på några av enkätens frågor. Redovisningen avslutas med rektors och lärares svar på några mer övergripande attitydpåståenden.

Vad anser rektorer och lärare om de nationella proven och det nuvarande nationella provsystemet?

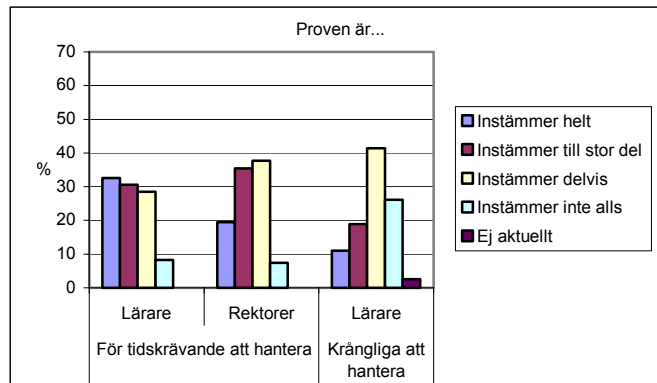
Proven möter skolan

De nationella proven innebär inte endast ett antal prov att genomföras av elever i år 5 eller i år 9. De medför även, för rektorer på skolnivå och för berörda lärare på klassrums- och lektionsnivå, en hel del insatser av logistiskt slag, termins-/lektionsplanering etc. för att passa in proven i skolans vardag.

Lärares och rektors enkätsvar visar att provverksamheten inte är utan kostnad tidsmässigt. Samtidigt uttrycker lärare och rektorer på intervjustolarna att proven är väl värda sina insatser, sett i förhållande till vad de ger tillbaka. Som vi ska se i kommande avsnitt uttrycker också enkätdata att proven värderas som viktiga av många lärare och rektorer.

Nästan två tredjedelar av lärargruppen och drygt halva gruppen rektorer anser att proven är för tidskrävande. Mindre än en av tio lärare och rektorer anser det motsatta. Färre lärare, ca tre av tio, tycker att de är krångliga att hantera (se figur 1).

Dnr 01-2003:2038



Figur 1. Lärares och rektors svar på frågan om proven är för tidskrävande att hantera och lärares svar på frågan om proven är krångliga att hantera.

Erfarna lärare, med många år i yrket, tycker inte att proven påverkar planeringen i någon högre utsträckning. Så är t.ex. 'nästan inte alls' ett tre gånger så vanligt svar bland de mest erfarna lärarna (med tio år eller fler i yrket) jämfört med nya lärare (med upp till två års erfarenhet). Att proven skulle vara för tidskrävande eller krångliga att hantera skiljer däremot inte i uppfattning mellan mer och mindre erfarna lärare.

Om provens (olika) syften och relativa värde

Uppfyller proven sina syften?

De nationella proven ska svara mot flera olika syften. Ställs frågan till rektorer och lärare i vilken grad man anser att proven uppfyller syftena, framträder en relativt tydlig bild. Rektorer anser i högre grad än lärare att proven faktiskt svarar mot det som de avses svara mot men både lärare och rektorer är i hög grad nöjda. Både rektorer och lärare värderar framförallt provens förmåga att stödja bedömning av måluppfyllelse, och som hjälp att identifiera elevers starka och svaga sidor (se tabell 5).

En något lägre andel rektorer och lärare anser att proven i hög eller mycket hög grad fungerar som stöd för en likvärdig bedömning/rättvis betygssättning, som ett förtydligande av det mål- och kunskapsrelaterade betygssystemet eller som förebildliga genom att peka på de kvaliteter som gäller för olika betyg.

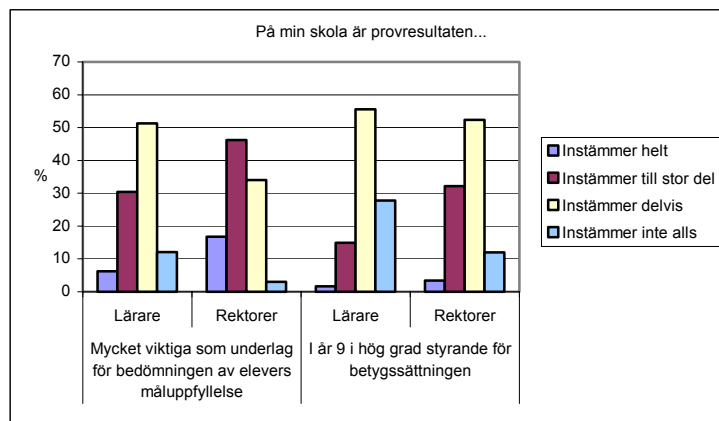
Mest tveksamma är lärarna och då tydligast om provens förebildliga syfte, att peka på olika betygskvaliteter eller som ett förtydligande av betygssystemet. Tre av tio lärare är inte nöjda.

Dnr 01-2003:2038

Tabell 5. Lärares och rektors svar på frågan "De nationella proven och de diagnostiska materialen är avsedda att fylla flera syften. I vilken grad anser du att de uppfyller följande syften?". Svarsfördelning i procent.

I vilken grad anser du att proven uppfyller följande syften?	I hög grad		I ganska hög grad		I ganska låg grad		Nästan inte alls	
	Lär	Rekt	Lär	Rekt	Lär	Rekt	Lär	Rekt
Som stöd för analys och bedömning av elevernas starka och svaga sidor	26	34	61	56	12	9	1	0
Som stöd i bedömningen av om eleverna har nått uppställda mål	26	40	62	53	11	7	1	0
Som stöd för principen om en likvärdig bedömning och en rättvis betygssättning	23	33	54	46	18	18	6	3
Som ett förtydligande av det mål- och kunskapsrelaterade betygssystemet	17	28	55	53	23	17	5	2
Som förebild genom att peka på de kvaliteter som gäller för olika betyg	18	28	53	49	22	19	8	4

Som figur 2 visar håller närmare två tredjedelar av rektorerna (helt eller till stor del) med om att provresultaten är mycket viktiga underlag för hur skolan bedömer elevernas måluppfyllelse. En dryg tredjedel av lärarna delar denna uppfattning (och nästan halva gruppen lärare i år 5). En dryg tredjedel av rektorerna jämfört med hälften så många, ca 15 procent, av lärarna anser proven i hög grad styrande för betygssättningen.



Figur 2. Lärares och rektors svar på frågan om proven är viktiga för bedömning respektive styrande för betygssättning.

Intervjuerna med lärare år 9 visar att det varierar hur lärare förhåller sig till relationen provresultat och slutbetyg i år 9, bilden är inte enhetlig. Några lärare framhåller att nationella proven ingår som endast en del i en samlad bedömning av elevers kunskaper i ämnet och inte heller kan förväntas svara mot hela årets samlade prestationer. "Det (nationella provet) är ju ändå bara en del. Det är ju hela året i nian som ska sättas betyg på". Andra lärare resonerar enligt någon princip, t.ex. att provresultat endast kan bidra till att höja, inte sänka, ett betyg. "Uppåt påverkar det mer. Blir en del i mina argument för ett högre betyg". Åter andra lärare säger att det förekommer att man sänker slutbetyget med hänvisning till provresultaten.

Dnr 01-2003:2038

Att proven fungerar som stöd för lärares bedömning framkommer mer generellt i intervjuerna. Det kan t.ex. handla om elever som ligger på gränsen till ett betyg där proven blir avgörande. ”Jag tycker man känner sig säkrare på att man sätter rätt betyg. Jag ser det som ett ganska stort stöd i betygssättningen.” Även om man tycker att man klarar av att bedöma elevernas kunskaper själv så fungerar ändå proven som en sorts koll på att man är på rätt spår, en bekräftelse. ”Det är väl mer av en bekräftelse att man ligger rätt, mer är det nog inte, kanske för några enskilda elever. Jag tycker nog att jag kan bedöma utan hjälp av nationella prov”. ”I stort sett bekräftar proven det man redan vet”.

Utöver att vara en bekräftelse på lärarens egen bedömning så framkommer också att proven ger ett stöd utåt, t.ex. i förhållande till besviktiga föräldrar. ”De nationella proven är en bra bekräftelse för mig, jag har torrt på fötterna och kan visa det här är inte jag som sätter för låga betyg. Det här året har jag en del föräldrar som har varit väldigt arga och tycker att det är skolans fel att deras barn inte klarar kraven. Då känns det skönt att veta att det inte är jag som har gjort de här proven.”

Ses proven som viktiga eller oviktiga?

Hur värderar man som rektor eller som lärare de nationella prov och diagnosmaterial som tas fram för deras bruk i skolan och i klassrummet, ses de som viktiga eller oviktiga?

De allra flesta lärare och rektorer ser ämnesproven som viktiga och återigen betonar rektorer i högre grad än lärare detta. Skillnaden är mindre uttalad för de diagnostiska materialen. Det är också färre bland de svarande totalt sett som anser dessa viktiga. Endast några få procent, såväl bland rektorer som bland lärare, anser det viktigt med en eventuell utbyggnad av provsystemet till att omfatta fler ämnen och fler år än nuvarande (se tabell 6).

Få lärare och rektorer anser att ämnesproven år 9 är oviktiga. Tvärtom betonar två tredjedelar av rektorerna och halva lärargruppen ämnesproven i år 9 som *mycket* viktiga. Ungefär samma svarsmönster gäller för proven i år 5, men färre anser dem ’mycket viktiga’. Företrädesvis är det lärare i år 9 som markerar år 5 proven som mycket viktiga, halva gruppen jämfört med en tredjedel av lärarna i år 5. Däremot skiljer det inte hur man som lärare i år 5 eller i år 9 värderar de nationella proven i år 9. Lika många lärare, oavsett i vilket år man undervisar, tycker att dessa prov är viktiga.

Dnr 01-2003:2038

Tabell 6. Lärares och rektorers svar på frågan "Hur viktigt är det att nationella prov och diagnosmaterial finns?" Svarsfördelning i procent.

Hur viktigt är det att det finns...?	Mycket viktigt		Ganska viktigt		Ganska oviktigt		Inte alls viktigt	
	Lär	Rekt	Lär	Rekt	Lär	Rekt	Lär	Rekt
Ämnesprov år 9	54	70	41	27	5	2	1	1
Ämnesprov år 5	45	61	45	34	8	4	2	1
Diagnosmaterial F-5	28	41	53	47	15	10	4	2
Diagnosmaterial 6-9	27	37	52	50	17	11	4	2
Nationella prov för andra år	9	7	20	17	44	51	28	25
Nationella prov i andra ämnen	7	7	20	21	44	44	29	29
Diagnostiska material i andra ämnen	9	11	32	29	37	39	23	21

En möjlig förklaring till att proven i år 5 ses som jämförelsevis mindre viktiga av lärare i år 5 än av lärare i år 9, är skillnaden i undervisning mellan lärargrupperna. Flertalet av dessa lärare i år 5 undervisar i fler (eller i alla) ämnen. För majoriteten av dessa lärare innebär det också att de ansvarar för de nationella proven i åtminstone två, ofta i alla tre ämnen. Detta innebär en större arbetsbörda för läraren som kan bidra till att man som lärare i år 5 är mindre positiv till proven. Möjligen handlar det också om att man arbetar med yngre elever. Proven innebär kanske därmed, för eleverna och för klassen, en relativt sett större påfrestning. Enbart mot bakgrund av enkätmaterialets de facto beskrivning är det inte möjligt att uttala sig om vad som döljer sig bakom denna skillnad i hur viktiga de nationella proven ses vara. Till viss del stöds dock resonemangen av de intervjuer som har genomförts med lärare i år 5.

Att proven innebär mycket arbete för lärarna i år 5 är något som bekräftas i intervjuerna. Den sammanlagda arbetsbördan är också något som lyfts fram som motiv för att inte vilja se en breddning av provsystemet till fler ämnen. Man menar också att eleverna redan är stressade över de prov som finns och att de pratar om vad som händer om man inte blir "godkänd". Att proven i år 5 ställer höga krav på organisering av verksamheten framkommer också i intervjuerna. Inte minst blir gruppuppgifter och muntliga prov stötestenar att komma runt. "Det var svårt även när vi hade assistent och var två stycken i klassen". Även om eleverna är vana vid gruppuppgifter blir lärarens roll vid genomförandet av ämnesprovet en annan, mer av att bedöma istället för att handleda och assistera.

För samtliga frågor om provens syften och hur viktiga proven är finns det en tendens till skillnader mellan lärare som svarar för olika ämnesprov. Skillnaderna är inte så markanta men mer genomgående. Lärare som svarar för provet i engelska år 9 är mer positiva till ett flertal av provens syften; bl.a. gällande principen om likvärdighet i bedömning och rättvis betygssättning och provens förebildlighet. De markerar också i högre grad hur viktigt ämnesprovet i år 9 är.

Dnr 01-2003:2038

En jämförelse mellan mer och mindre erfarna lärare visar också en tendens till skillnader. Gruppen minst erfarna lärare har en avvikande profil. De är mer tveksamma till om proven verkligen uppfyller syftet att verka för likvärdighet och rättvisa. De är också mer tveksamma vad gäller provens förmåga att verka förebildliga, såväl vad avser det mål- och kunskapsrelaterade betygssystemet som provens förmåga att illustrera vad som gäller för olika betygskvaliteter.

Mindre erfarna lärare är också mindre övertygade om ämnesprovets värde, färre ser dem som mycket viktiga. Det gäller både ämnesproven i år 5 och i år 9. Däremot finns det inga skillnader i hur dessa lärare, jämfört med resten av lärargruppen, värderar de diagnostiska materialen. Möjligtvis finns det snarast en tendens till att de är mer positiva till dessa material. Denna tendens går också igen hur de värderar ett breddande av provsystemet till fler ämnen eller år. Här tycker de i likhet med resten av gruppen lärare att det inte är särskilt angeläget med fler ämnesprov, vare sig för andra åldersgrupper eller för andra ämnen. Däremot tycker fler bland dessa "yngre" lärare att det är angeläget med diagnostiska material i fler ämnen.

I intervjuerna med lärarna framhålls provens kvalitet, och här skymtar också provens legitimitet i verksamheten. "Trots att det är mer jobb (med proven) tycker jag verkligen att det är väl värt det. För jag kan inte göra så här bra prov, det tar jättelång tid att sätta ihop. Det märker jag i andra ämnen när jag försöker göra bra prov där". Även bland lärare i år 5 framkommer en generellt positiv värdering av proven, trots den tid och arbetsinsats de kräver.

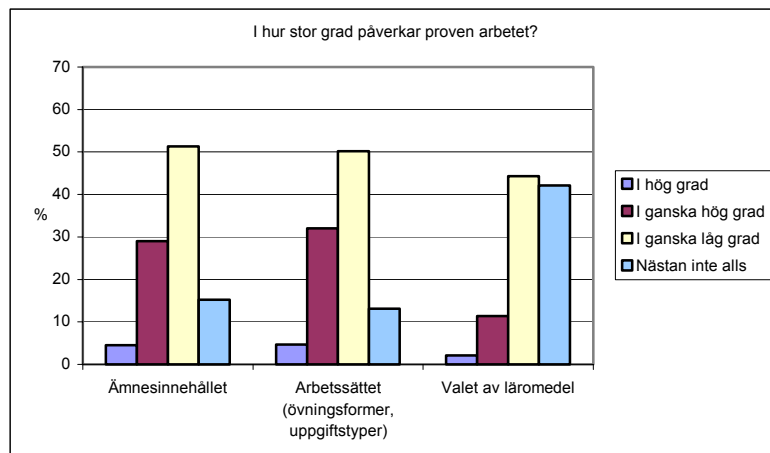
Hur påverkar proven skolans arbete och undervisning och hur stämmer de med lokala mål och kriterier?

Givet ett decentraliserat styrsystem med inbyggda frihetsgrader för kommuner, skolor och lärare att forma sin verksamhet kan man fundera över hur de nationella proven kommer in. I vilken grad uppfattar användarna t.ex. att proven styr undervisningens innehåll och val av metoder?

Påverkar/styr proven undervisningen?

Flertalet lärare anser inte att proven påverkar undervisningen i någon större utsträckning. Ungefär en tredjedel av lärarna menar dock att proven har denna effekt, både för val av ämnesinnehåll som för arbetssätt (övningsformer, uppgiftstyper). Än färre lärare tycker att proven påverkar valet av läromedel i någon högre grad, ungefär 15 procent. (se figur 3).

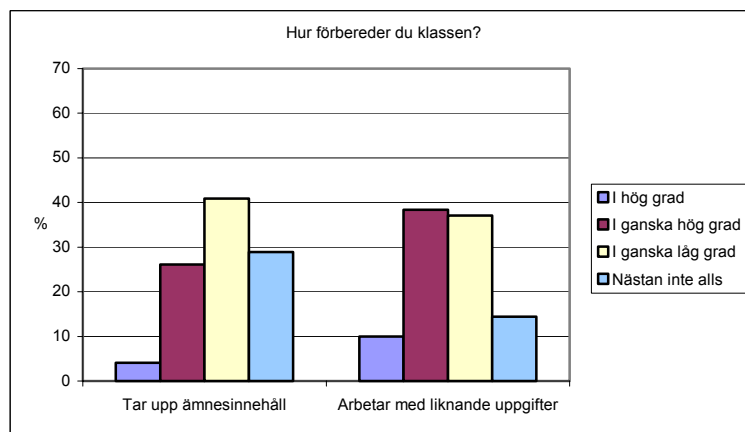
Dnr 01-2003:2038



Figur 3. Lärares svar på frågan om proven påverkar arbetet.

Det är inte heller självklart att påverkan ses som negativt, eller positivt. Läger man ihop lärarnas erfarenhet med deras värdering av provens påverkan framtonar en bild av lärarnas möjlighet att själva avgöra hur mycket proven ska få påverka undervisningen. Det finns ett tydligt samband mellan hur man värderar denna påverkan, som antingen negativ eller positiv, och hur man beskriver att proven faktiskt påverkar eller inte påverkar undervisningen. De lärare som ser påverkan som negativ är också till övervägande del samma lärare som beskriver att proven de facto inte påverkar undervisningen nämnvärt. På samma vis ses inte påverkan som negativt bland majoriteten av lärare som beskriver att proven faktiskt påverkar.

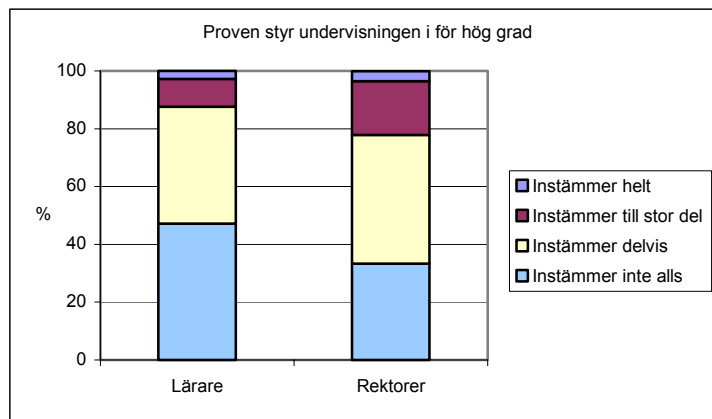
Flertalet lärare beskriver också att man inte anpassar undervisningen särskilt till proven. Ungefär sju av tio lärare, oavsett undervisning i år 5 eller år 9, utgår inte från provens ämnesinnehåll i utformandet av undervisningen. Det är vanligare att låta eleverna arbeta med uppgifter som liknar de som förekommer i de nationella proven. Det menar ungefär halva gruppen lärare och särskilt lärare i år 9 (se figur 4).



Figur 4. Lärares svar på frågan om hur de förbereder klassen på olika sätt (år 9).

Dnr 01-2003:2038

En majoritet av lärare tycker inte att proven styr undervisningen i för hög grad. Fler rektorer, och framförallt rektorer med ansvar för år 9, ser dock proven som i för hög grad styrande. Andelen uppgår till en dryg femtedel jämfört med en dryg tiondel av lärarna (se figur 5).



Figur 5. Vad anser du om proven? Uppdelat på lärare och rektorer.

Det skiljer också mellan lärare med mer och mindre erfarenhet. Drygt hälften av lärarna med fler än tio år i yrket instämmer inte alls i påståendet att proven skulle styra undervisningen i för hög grad. Det kan jämföras med en tredjedel av lärare med färre än två år i yrket. Enligt samma mönster är det fler ”nya” lärare som anser att proven påverkar undervisningen, både val av innehåll och arbetssätt jämfört med ”gamla” lärare. Det är också väsentligt fler av ”nya” lärare som förbereder sin klass särskilt inför de nationella proven, både avseende val av ämnesinnehåll och val av uppgifter. Till exempel menar en majoritet, knappt två tredjedelar, av mindre erfarna lärare att man förbereder eleverna inför proven genom att arbeta med liknande uppgifter. Motsvarande andel bland lärare med mer erfarenhet är lägre, drygt en tredjedel.

I intervjuerna framkommer för några lärare att proven inte styr undervisningen så mycket som bedömningen. ”Handlar det om ett friare ämne har man kanske oftare vidare ramar för vad man godkänner, man tillåter mer. I vanliga fall resonerar jag så att det inte är ämnet (för uppgiften) som är viktigt utan det handlar mer om hur resultatet blev, mindre viktigt att man har löst själva uppgiften. I proven blir det tydligare att se att man faktiskt har en uppgift och visa att man har läst instruktionerna och följt dem”.

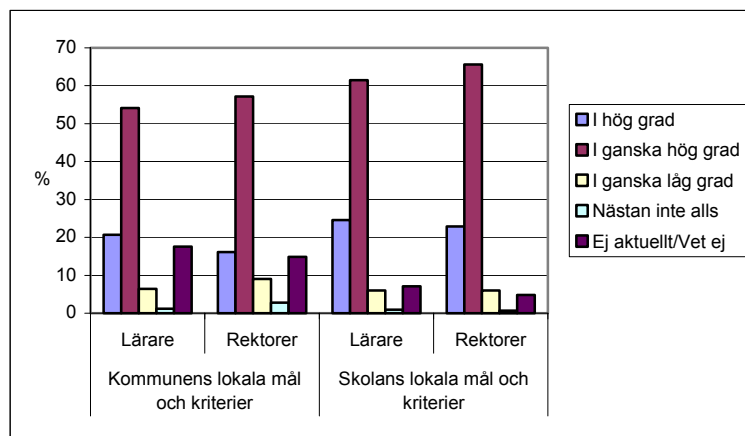
Stämmer proven överens med lokala mål?

Endast några få procent av lärarna och rektorerna anser att de kunskaper som proven bedömer avviker från skolans respektive den kommunala nivåns uppställda mål (se figur 6). Ungefär tre fjärdedelar av rektorer och lärare tycker att proven stämmer i ganska eller i

Dnr 01-2003:2038

hög grad överens med kommunens lokala mål och kriterier. Ännu fler, ca 85 procent, tycker att de stämmer med skolans lokala mål och kriterier. En relativt stor grupp av både lärare och rektorer, ca 15 procent, känner inte till vad kommunen har för lokala mål och kriterier. Och inte fullt lika många men ändå sex, sju procent av lärare och rektorer känner inte till den egna skolans mål och kriterier (?!).

Att så många uppger sig inte känna till kommunens lokala mål och kriterier kan delvis bero på att frågans formulering är otydlig och att det inte är självklart för de svarande vad som avses.



Figur 6. I vilken grad stämmer de kunskaper som bedöms i proven överens med...? Uppdelat på lärare och rektorer.

På frågan om de nationella proven minskar behovet av att arbeta med kursplaner, konkretisering av betygskriterier eller att använda lokala prov, menar majoriteten av lärare och rektorer att proven inte har dessa effekter (se tabell 7). Allra minst påverkat är arbetet med kursplaner. Drygt två tredjedelar av gruppen håller inte alls med om att detta arbete minskar. Att behovet att konkretisera betygskriterier eller att använda lokala prov minskar i och med de nationella proven håller majoriteten av lärare och rektorer inte heller med om.

Tabell 7. Lärares och rektors svar på frågan "Vilka effekter har de nationella proven för följande?" Svarsfördelning i procent.

Proven minskar i hög grad behovet av...	Instämmer helt		Instämmer till stor del		Instämmer delvis		Instämmer inte alls	
	Lär	Rekt	Lär	Rekt	Lär	Rekt	Lär	Rekt
Att arbeta med kursplanerna	1	1	6	7	26	20	68	72
Att konkretisera betygskriterierna	1	2	11	12	33	28	55	58
Att använda lokala prov	4	3	11	14	29	39	56	44

I dessa avseenden gällande provens överensstämmelse med den lokala nivån finns det en tendens till skillnader mellan lärare som svarar för olika ämnesprov. Skillnaden finns återigen mellan lärare som undervisar i år 9 i engelska respektive i matematik. Lärare i engelska menar i högre grad att provens innehåll stämmer såväl i förhållande till skolans mål som till den kommunala nivåns mål. Den motsatta

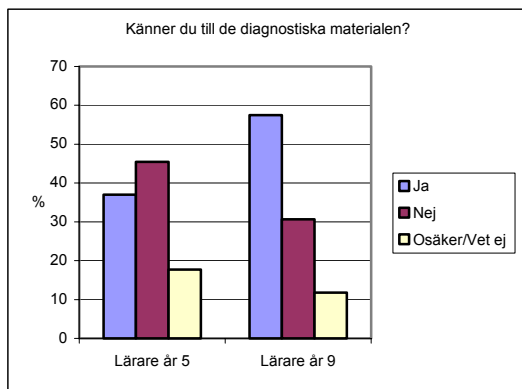
Dnr 01-2003:2038

tendensen är tydlig hos lärare i matematik. Relativt sett tycker färre matematiklärare att proven stämmer med såväl den lokala som den kommunala nivåns mål. Väsentligt fler lärare i matematik tar t.ex. helt avstånd ifrån antagandet att behovet att använda lokala prov skulle minska i och med förekomsten av nationella prov i ämnet (sju av tio jämfört med ungefär fem av tio lärare om proven i engelska och svenska).

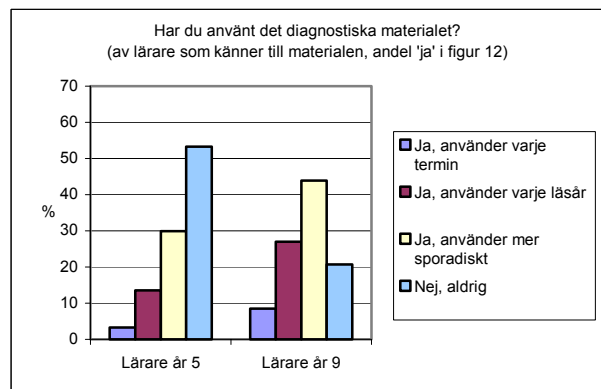
Hur används de nationella diagnostiska materialen, om de överhuvudtaget används?

Jämfört med ämnesproven värderas de diagnostiska materialen både av rektorer och av lärare som mindre viktiga. Generellt är det fler rektorer än lärare som värderar materialen som *mycket* viktiga, framförallt diagnosmaterialet upp till år 5. Hur ser det då ut med användningen på skolan?

Inledningsvis ställdes en fråga om man som lärare kände till Skolverkets diagnostiska material i svenska, engelska och matematik. Drygt hälften av lärargruppen visade sig ha kännedom om materialen, resten var antingen osäkra eller svarade nej på frågan (se figur 7). Av lärare med kännedom om materialen är det ungefär tre av tio som regelbundet använder sig av dem (se figur 8). Regelbundet innebär för de flesta ett användande varje läsår. En mindre grupp använder materialen varje termin. En dryg fjärdedel av lärarna totalt, och hälften av lärarna i år 5, använder sig inte alls av de diagnostiska materialen.



Figur 7. Andel lärare år 5 och år 9 som känner till de diagnostiska materialen.

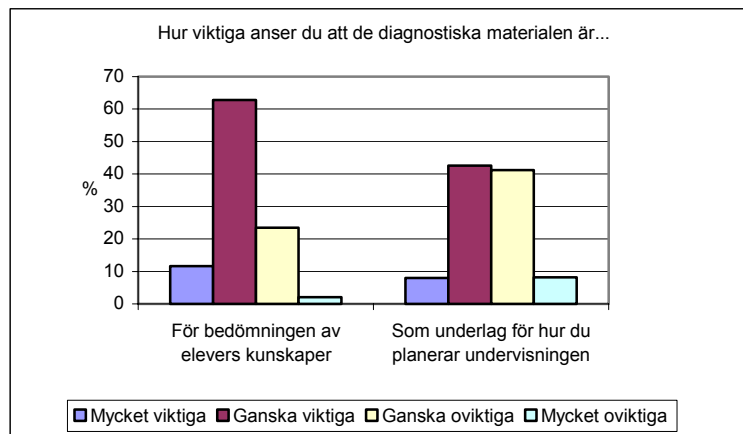


Figur 8. Andel lärare år 5 och år 9 som använt det diagnostiska materialet (av lärare med kännedom om materialen).

Materialen används vanligen för (nästan) samtliga elever. Sex av tio lärare redovisar detta, medan resten antingen använder materialet i grupper av elever eller för enskilda elever som man känner sig osäker på. Att använda materialen för enskilda elever är mer förekommande bland lärare i år 5. Tre av fyra lärare anser att materialen är viktiga för bedömning av elevers kunskaper. Drygt hälften anser dem viktiga som underlag för planering av undervisning (se figur 9). Möjligtvis

Dnr 01-2003:2038

kan man notera att det bland lärare som ändå väljer att använda materialen är förhållandevis många som värderar deras betydelse som ganska oviktig. Framförallt gäller detta materialens betydelse som underlag för hur man planerar sin undervisning men också dess betydelse som underlag för bedömningen av elevers kunskaper. Hur man värderar de diagnostiska materialens betydelse skiljer inte nämnvärt mellan lärare som undervisar i år 5 eller i år 9.



Figur 9. Lärare om hur viktiga de diagnostiska materialen är.

Riktat man istället uppmärksamheten mot de lärare som inte använder materialen, vad anges då för skäl? Den vanligaste orsaken att inte använda Skolverkets diagnostiska material är att man istället använder andra diagnostiska material. Ett annat skäl, för en mindre grupp, är att man inte upplever sig ha behov av sådana hjälpmedel.

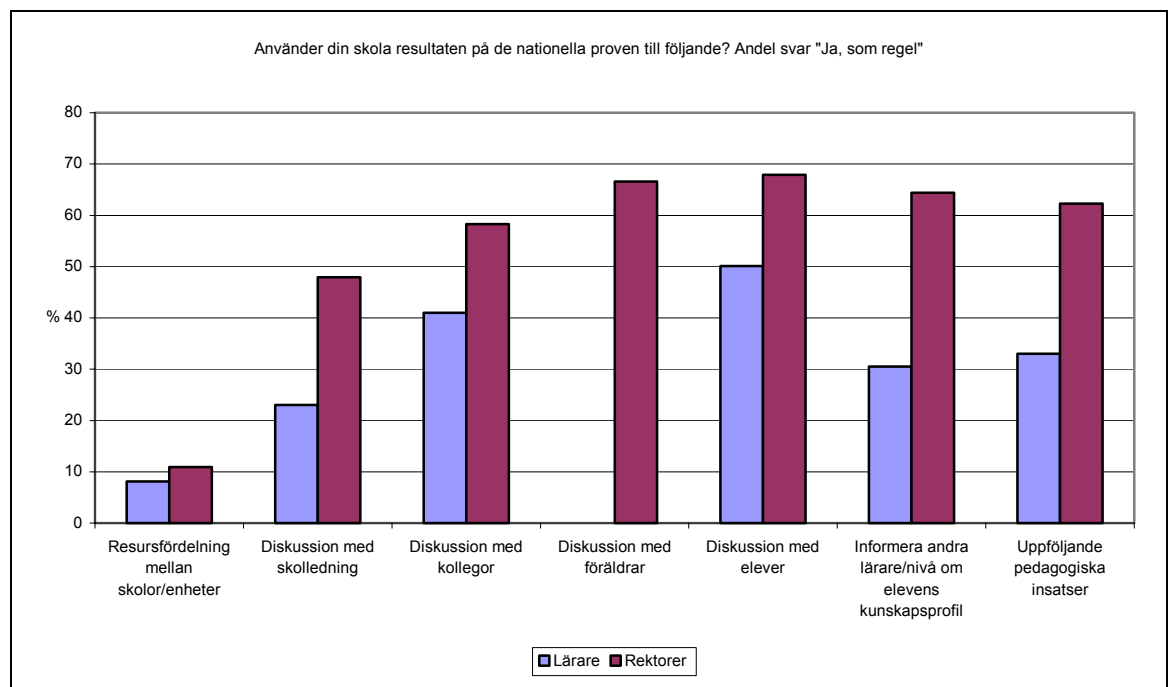
Ser man till hur rektorerna har beskrivit användningen av dessa material indikerar deras svar en mer utbredd användning. Nästan åtta av tio rektorer menar att skolan använder sig av materialen. Att notera är att medan rektorns svar representerar hela skolan representerar lärarsvaren enbart år 5 och år 9. Detta kan i sig betyda att materialen också används av andra lärare och för elever i andra år på skolan. En ytterligare tolkning är att lärarsvaren ligger närmare "sanningen" och att rektor överskattar användningen av materialen, på avstånd från den faktiska undervisningspraktiken. Ungefär en av tio rektorer uppger sig inte känna till materialen. Som anledning till varför man inte använder Skolverkets material uppger även rektorerna oftast att man istället använder andra diagnostiska material.

Vanligaste användningen av materialen är i diskussioner med antingen elever eller kollegor. Lite mindre vanligt är det att använda materialen i diskussion med föräldrar och ännu mer sällan används de i diskussion med skolledningen. Företrädesvis är det lärare i år 5 som använder materialen i diskussion med elever och föräldrar. En fjärdedel av lärarna i år 9 använder inte materialen alls för diskussion med föräldrarna. Fyra av tio lärare i år 9 använder inte heller materialen för diskussion med skolledning.

Dnr 01-2003:2038

Hur använder skolan sig av provresultaten - förmedling, samverkan, samarbete?

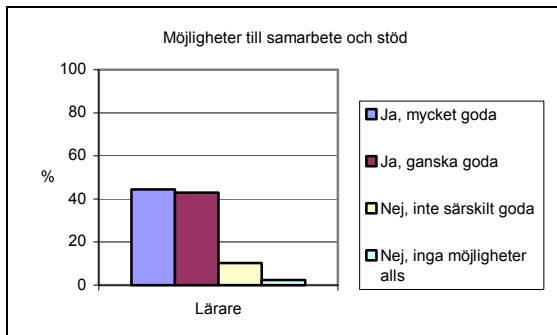
Majoriteten av lärare och rektorer, varierande mellan sjuttio och nitio procent, uppger att de använder sig av provresultaten för diskussion med såväl elever, föräldrar, kollegor som med skolledning, oftast med elever och kollegor (se figur 10). Att provresultaten används regelmässigt för uppföljande pedagogiska insatser alternativt som information till andra lärare/högre nivåer menar sex av tio rektorer och lika många lärare i år 5. Att färre lärare i år 9 anger dessa alternativ som användning av provresultat är inte oväntat med tanke på att dessa prov infaller först under grundskolelevens sista år och termin. Betyddigt ovanligare är det då att använda sig av provresultaten som underlag för diskussioner om resursfördelning mellan skolor/enheter, endast en av tio rektorer och lärare uppger 'ja, som regel' och två av tio uppger 'ja, ibland'.



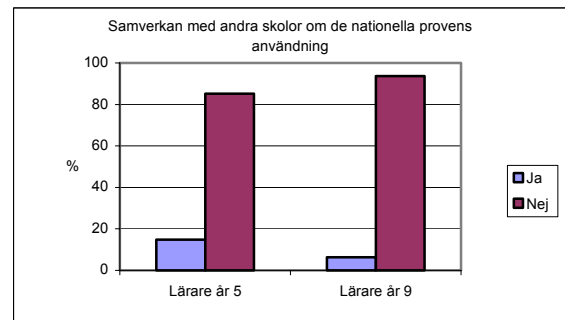
Figur 10. Andel lärare och rektorer som svarar "ja, som regel" (svarsalternativ "diskussion med föräldrar" föll bort ur lärarenkäten och finns därför bara för rektorer).

En majoritet, åtta av tio lärare, uppger sig ha ganska eller mycket goda möjligheter till samarbete och stöd bland sina kollegor om de nationella provens användning (se figur 11). Bilden är den rakt motsatta vad gäller om skolan samverkar med andra skolor om provens användning (se figur 12). Nio av tio uppger att man inte samverkar. Det är framförallt lärare i år 5 som uppger samverkan i de få fall som detta uppges.

Dnr 01-2003:2038



Figur 11. Finns det goda möjligheter till samarbete och stöd bland dina kollegor vad gäller de nationella provens användning, t.ex. för bedömning/betygssättning av elever? (Endast lärare).

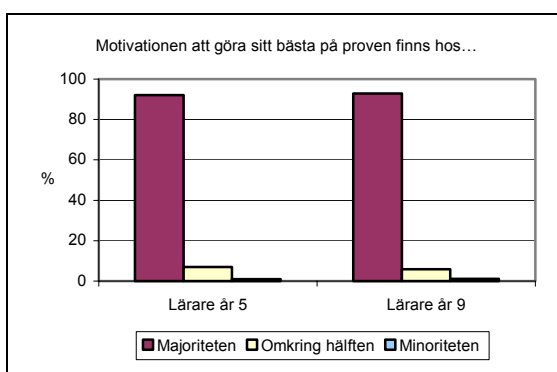


Figur 12. Samverkar man på din skola med andra skolor om de nationella provens användning? Uppdelat på lärare i år 5 och år 9.

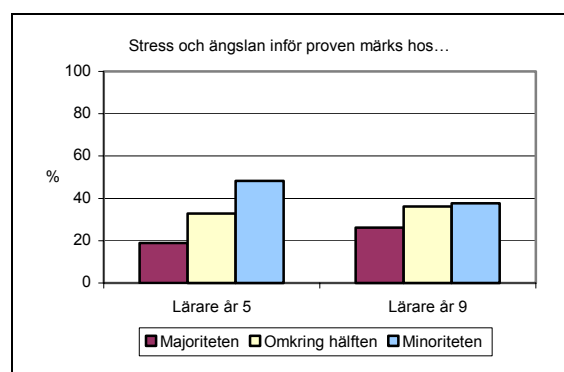
Vad säger lärarna om elevers reaktioner inför proven och vad säger eleverna om proven?

Ett antal frågor i lärarenkäten handlade om elevernas reaktioner inför och efter provens genomförande. Vilka reaktioner, positiva eller negativa, möter lärare från eleverna?

Nio av tio lärare beskriver sin klass som ambitiös i meningen att majoriteten av eleverna är motiverade att göra sitt bästa på de nationella proven (se figur 13). Samtidigt är det en fjärdedel av lärarna i år 9 och något färre, en femtedel av lärarna i år 5 som uppger att en majoritet av eleverna också är stressade och ängsliga inför ett prov (se figur 14). Ungefär en fjärdedel av lärarna instämmer i påståendet ”att proven innebär ett för stort stressmoment för eleverna”.



Figur 13. Hur beskriver du klassens reaktioner inför proven? Lärare i år 5 och år 9.

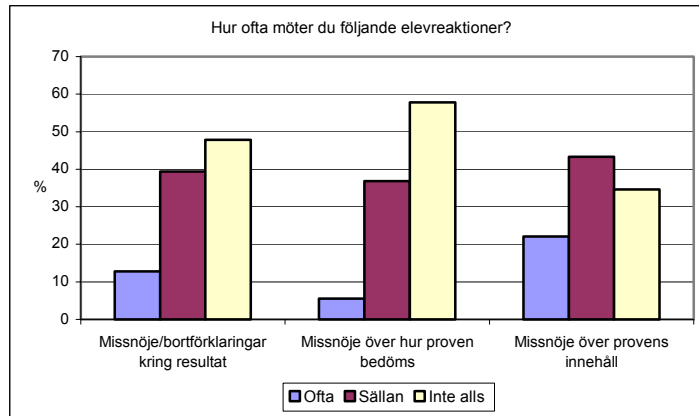


Figur 14. Hur beskriver du klassens reaktioner inför proven? Lärare i år 5 och år 9.

Rena missnöjesyttringar från eleverna över antingen provens innehåll eller hur de bedöms är inte vanligt förekommande enligt lärarna (se figur 15). Att elever uttrycker sitt missnöje genom ”bortförklaringar” av resultaten förekommer inte heller i någon högre utsträckning. Generellt sett är denna typ av elevreaktioner mindre vanliga, men som

Dnr 01-2003:2038

man möjligtvis kan förvänta sig är de i den mån att de ändå förekommer genomgående vanligare i klassrummen i år 9 än i år 5.

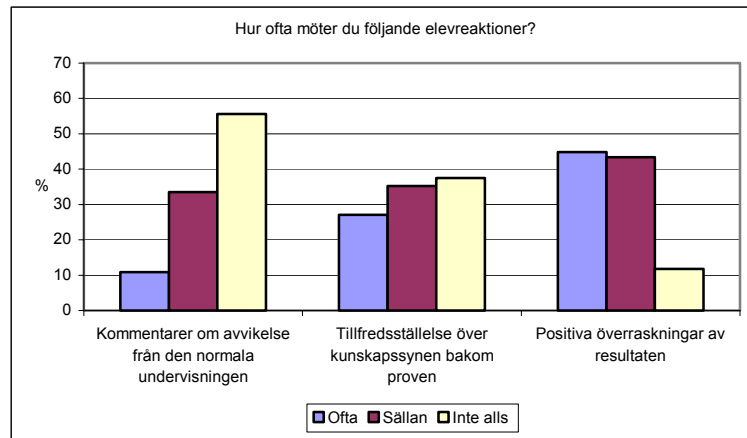


Figur 15. Lärares erfarenhet av olika elevreaktioner på proven.

Samtalar man med eleverna om vad de tror proven spelar för roll för betygsättningen så finns det en osäkerhet och oklarhet. Här finns det också olika åsikter. Eleverna tar upp att det skiljer mycket mellan ämnen, men också mellan olika lärare. Eleverna tror ändå att provet bidrar till likvärdighet: ”Jo, det är bra för att veta att alla skolor lär ut lika mycket, att inte skolorna är på olika nivå eller så”. Eleverna har också en del andra sätt att resonera kring betygsnivåer och betygsättning: ”Man känner ju ungefär vilket betyg man har jämfört med hur bra man är med resten av klassen och hur bra läraren verka tycka att man är, så man märker ju liksom vilket betyg man ska få, oavsett vad som står i några betygsriterier.” (I början av terminen presenteras betygsriterierna för eleverna, men det verkar som det tas in mer ytligt och sedan helt glöms bort.)

Hur vanligt är då mer positiva elevreaktioner på proven, i den mån de förekommer? Hälften av lärarna som undervisar i år 5 möter ofta elever som är nöjda med sina resultat. Fyra av tio lärare i år 9 beskriver samma sak. Det är inte heller ovanligt enligt lärarna att möta reaktioner som vittnar om att eleverna är positiva till provens innehåll, till den kunskapssyn som ligger bakom proven (se figur 16). Detta uttrycker en tredjedel av lärare i år 5 och en fjärdedel av lärare i år 9. För år 9-lärarna är det jämförelsevis vanligare att möta nöjda elever i engelskan. Även om det inte är ovanligt med positiva reaktioner på proven är det ändå en majoritet av lärargruppen som menar att deras elever mycket sällan eller inte alls uttrycker sig positivt.

Dnr 01-2003:2038



Figur 16. Lärares erfarenhet av olika elevreaktioner på proven.

Samtalen med elever visar en rätt varierad bild av uppfattningar om proven. Någon grupp elever är betydligt mer negativa och uppfattar t.ex. proven som mycket svårare än de vanliga proven eller gamla prov som man har fått öva på innan (elever om ma-prov): "Mycket svårare, man ska ju fatta! Det finns ju inga förklaringar i våra böcker."

Andra elever är mer positiva i sina beskrivningar om provet. Det har också en betydelse, tas mer på allvar, att provet är nationellt. "Det känns större och viktigare eftersom de ska in på något lager och är sekretessbelagt. Det är mycket betygen i det också, jag tror det." Det finns också en känsla av att dessa prov är mer rättvisa än andra prov. "På det nationella visar man vad man kan, annars är det bara att träna in till det" (Här känner man inte igen oförberedda prov som del av vanliga undervisningen).

Från intervjuerna med lärare så framtonar en bild av att elever uppskattar proven. "Jag menar att de har varit positiva, och till alla tre (delprov), det är ju faktiskt tre ganska roliga prov att göra". Nervositet finns också med, men inte nödvändigtvis negativt. "Det är ju också lite grann en känsla av att de gör något som, och det är lite häftigt, alla gör samtidigt i hela landet och att det är någonting lite större. Det tror jag de tycker är kul".

Som vi har sett tidigare så menar majoriteten av lärare att man inte anpassar sin undervisning i någon högre grad efter innehåll i proven. Hur vanligt förekommande är det då att eleverna kommenterar att proven avviker i förhållande till den normala undervisningen?

Inte så många, men ungefär var tionde lärare i år 5 eller år 9 anger att kommentarer om avvikelse mellan prov och undervisning förekommer frekvent (mycket eller ganska ofta). Som lärare i matematik i år 9 träffar man jämförelsevis oftare på elever som tycker att proven avviker i förhållande till vad, eller kanske hur, de lär matematik på lektionerna. Tre av tio lärare i svenska eller i engelska har aldrig hört denna kommentar om provet. Motsvarande andel är inte fullt en av

Dnr 01-2003:2038

tio lärare i matematik. Tydligast är skillnaden i år 9 men är också märkbar i lärares beskrivning som undervisar i år 5.

Fungerar proven på samma sätt för olika elevgrupper eller finns det elever som proven systematiskt missgynnar i deras chanser att lyckas på proven? Vad säger lärarna?

På frågan om proven i sig själva (dvs. inte ämnet) ”slår” olika för pojkar eller flickor, menar tre av fyra lärare att det inte gör det. För den tiondel av lärarna som menar att proven gynnar flickor är detta svar något vanligare bland lärare som svarar för provet i svenska (både år 5 och år 9). En femtedel av lärarna anger att provet är diskriminerande, men åt bägge hållen beroende på delprov.

På frågan om provens innehåll/utseende på något sätt missgynnar elever med annat modersmål än svenska uppger en majoritet av lärare att de inte vet eller kan ta ställning. Men ungefär en tiondel av undervisande lärare i år 5, och något fler lärare i år 9, menar att proven missgynnar dessa elever. Bland lärare i år 9 är det vanligare att lärare som svarar för matematikprovet uppger att provet är diskriminerande i detta avseende. Bland lärare i år 5 finns samma tendens att fler matematiklärare, men också något fler lärare som svarar för svenskprovet, anser att provet missgynnar denna grupp elever.

Med tanke på att det är så stor grupp lärare som uppger att de inte kan besvara frågan, om proven missgynnar elever som har annat modersmål än svenska, är det motiverat att se närmare på den grupp lärare som enligt enkäten undervisar i klasser med många av dessa elever. Hur ser de på provet i detta avseende?

Betydligt fler lärare som undervisar i klasser med minst en fjärdedel elever med annat modersmål instämmer att proven faktiskt missgynnar denna grupp. För proven i svenska och i matematik är denna uppfattning tre gånger så vanlig. Inte så oväntat är det färre lärare som bedömer provet i engelska som missgynnande. Kanske är det mer oväntat att andel lärare som svarar att matematikprovet missgynnar är så stor, t o m en något högre andel än lärare som menar det om provet i svenska (fyra av tio jämfört med tre av tio)?

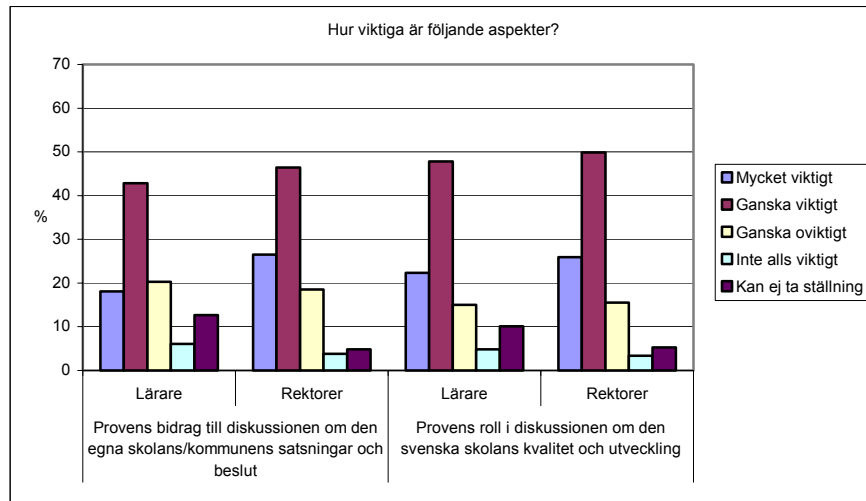
Lärares och rektorers uppfattning om proven och provsystemet – några övergripande attitydfrågor i enkäten

Den överlag positiva beskrivning som framtonat av lärares, rektorers och till viss del även elevers syn på de nationella proven och provsystemet bekräftas i några av enkätens mer generella påståenden.

Här ställdes t.ex. frågan om hur man som lärare eller rektor värderade provens bidrag till diskussionen om den egna skolans/kommunens beslut, och vidare, om provens roll i diskussionen om den svenska skolans kvalitet och utveckling. En majoritet av lärare och ännu fler rektorer menar att proven är viktiga ur dessa avseenden. Nästan två tredjedelar anser att provresultaten är viktiga för den lokala diskus-

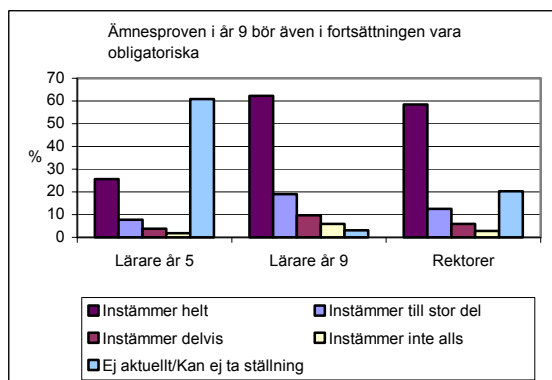
Dnr 01-2003:2038

sionen om satsningar och beslut. Ännu fler, sju av tio, anser dem viktiga vad gäller den generella diskussionen om svenska skolans kvalitet (se figur 17).

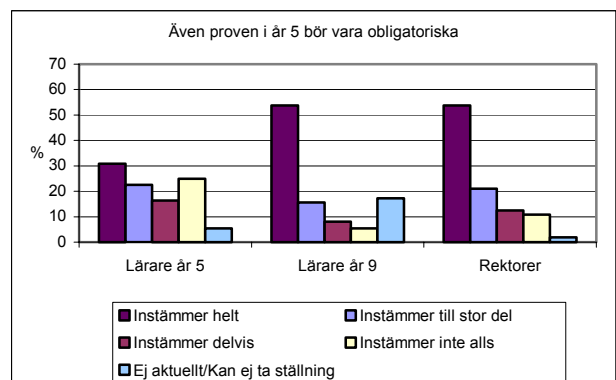


Figur 17. Lärares och rektors svar på frågor om provens bidrag i diskussionen om skolan.

På en mer konkret nivå ställdes några påståenden om det nuvarande provsystemets utformning, bl.a. om proven ska vara obligatoriska eller inte. Om ämnesproven i år 9 anser sju av tio lärare och rektorer att de även i fortsättningen ska vara obligatoriska (se figur 18). Drygt hälften av de svarande, och sex av tio lärare i år 9, uttrycker detta otvetydigt med ett 'instämmer helt'. På frågan om även ämnesproven i år 5 borde vara obligatoriska är det nästan samma andel, sju av tio, som ställer sig positiva till detta (se figur 19). Och återigen är det i det närmaste hälften som instämmer helt. Inte lika många lärare som undervisar i år 5 är däremot positiva till att göra dessa prov obligatoriska. En fjärdedel av lärare i år 5, jämfört med fem-tio procent av lärare i år 9 eller rektorer, kan inte alls tänka sig denna utveckling.



Figur 18. Hur väl stämmer följande påståenden med din erfarenhet/din uppfattning?

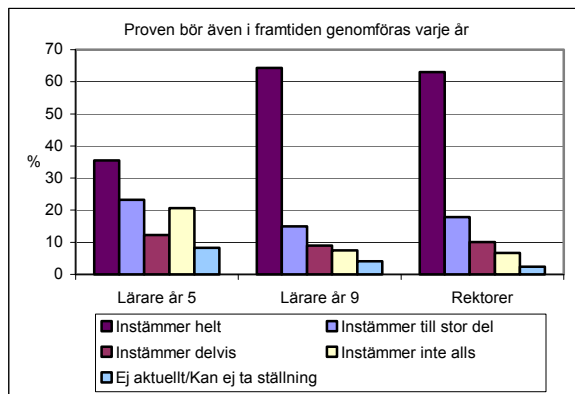


Figur 19. Hur väl stämmer följande påståenden med din erfarenhet/din uppfattning?

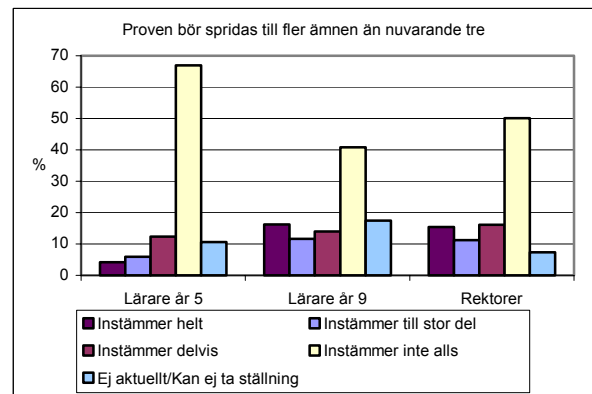
Det är också väsentligt färre lärare i år 5 som tycker att proven även fortsättningsvis bör genomföras varje år (se figur 20). Att proven

Dnr 01-2003:2038

skulle spridas till fler ämnen än nuvarande tre har lärare och rektorer uttryckt som oviktigt. Även här är det halva gruppen som helt tar avstånd ifrån en sådan breddning av provsystemet. Återigen är fler lärare i år 5 tveksamma, drygt två tredjedelar vill inte alls se nationella prov i fler ämnen (se figur 21). Enligt samma mönster är det också en majoritet, tre av fyra, av de svarande som inte håller med om att nuvarande provsystem skulle medföra en alltför ensidig fokusering på svenska, engelska och matematik. Fler rektorer än lärare ser dock denna risk.



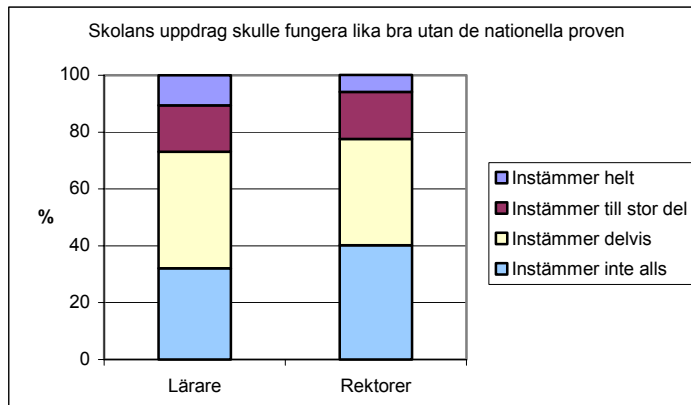
Figur 20. Hur väl stämmer följande påståenden med din erfarenhet/din uppfattning?



Figur 21. Hur väl stämmer följande påståenden med din erfarenhet/din uppfattning?

Ett lite mer provocativt påstående får avsluta redovisningen av lärares och rektors uppfattning om proven i grundskolan: "Skolans uppdrag skulle fungera lika bra utan de nationella proven". En tredjedel av lärare och rektorer håller inte alls med och lägger man till den andel som uttrycker ett 'instämmer delvis' så är det ungefär sju av tio som åtminstone inte oreserverat instämmer i påståendet (se figur 22). Här kan noteras att det är färre som tar helt avstånd bland lärare i år 5, en femtedel jämfört med 35-40 procent av övriga lärare och rektorer. Man kan också notera att detta ändå innebär att ungefär en fjärdedel av svarande håller med om att skolan kan klara uppdraget lika bra utan nationella prov, något ovanligare bland rektorerna och mer vanligt för lärare i år 5.

Dnr 01-2003:2038



Figur 22. De nationella proven sägs ibland ha vissa negativa effekter. Vad anser du om proven?

Några huvudsakliga resultat och preliminära slutsatser – en sammanfattning

Kapitlet har syftat att ge ett underlag för diskussionen om det nationella provsystemets kvalitet och effektivitet utifrån ett användarperspektiv. Röster från skolfältet har fått komma till tals och säga sitt om de nationella provens syften, relativa värde och användning. Rektorer, lärare i nationella provämnen i år 5 och år 9 och några grupper av elever har alla gett sin bild av hur proven ”passar in”; i undervisningen och mer generellt i skolans verksamhet. Hur resultaten används, diskuteras och hur man ser på en eventuell breddning av provsystemet till att omfatta fler ämnen och andra åldersgrupper har också belysts.

En generell slutsats är att de nationella proven förefaller äga legitimitet i systemet. Detta har bl.a. att göra med hur väl proven ses svara mot sina syften. Här anser en majoritet av både lärare och rektorer att proven väl uppfyller sina syften, framförallt som stöd för bedömning av måluppfyllelse och som hjälp att identifiera elevers starka och svaga sidor.

Bilden bekräftas i den allmänt positiva värderingen av hur viktiga proven ses vara. Trots att många lärare (och en del rektorer) beskriver att proven är allmänt resurskrävande (tid, planering etc.) menar man ändå generellt att proven är väl värda dessa ansträngningar, sett i relation till vad de ger tillbaka. Nio av tio lärare och rektorer anser att proven är mycket eller ganska viktiga. Den positiva bilden balanseras dock något av att en av fyra lärare och rektorer ändå håller med om att skolans uppdrag kan klaras lika bra utan nationella prov.

En annan generell slutsats är att man, som lärare och som rektor, i samma höga utsträckning är ointresserad av en eventuell breddning

Dnr 01-2003:2038

av provsystemet, vare sig det handlar om att innefatta fler ämnen eller andra åldersgrupper elever. Däremot är flertalet lärare och rektorer positiva till den nuvarande utformningen, som t.ex. med obligatoriska prov som återkommer varje år.

Ett genomgående mönster i denna värdering är att rektorerna generellt är mer uttryckligt positiva om (de existerande) proven och deras relativa värde. Detta kan ha att göra med flera förhållanden. Ett sådant förhållande är att man som rektor dels sitter längre ifrån alla praktiska handhavanden som proven kräver och också på mer avstånd ifrån bekymret att värdera provresultat i bedömning av elevers kunskaper. Ett annat förhållande som kan antas spela in är att rektorerna i första hand kan sägas representera skolnivån där provresultat på en aggregerad nivå har kommit att få en allt större betydelse. Detta är synligt i enkätens resultat där en majoritet av rektorer, men också många lärare, ser provens roll som viktig; i diskussionen om den lokala skolans villkor och förutsättningar och i diskussionen om den svenska skolans kvalitet och utveckling.

Ser man till lärarnas svar om utformandet av undervisningen förefaller proven inte påverka vare sig ämnesinnehåll, arbetssätt eller val av läromedel i någon högre utsträckning. Man anpassar inte heller undervisningen särskilt till proven. Detta kan också relateras till att flertalet lärare inte anser att proven är avvikande i förhållande till lokala mål och kriterier. Dock skiljer det i dessa avseenden mellan lärare med olika lång yrkeserfarenhet. Mindre yrkeserfarna lärare tycker i högre grad att proven inverkar på deras undervisning. Det är också väsentligt färre erfarna lärare som håller med om att proven skulle vara för styrande för undervisningen. Här kan man tillägga att de lärare som beskriver att proven påverkar undervisningen i någon högre utsträckning oftast värderar detta positivt. En möjlig tolkning är att man ser proven som stödjande snarare än som styrande.

De allra flesta lärare beskriver huvuddelen av elever i klassen som motiverade att göra sitt bästa på proven. Samtidigt pekar ganska många lärare på att en majoritet elever är stressade och ängsliga inför proven. Var fjärde lärare tycker att proven innebär ett för stort stressmoment för eleverna. Att proven fungerar olika för olika grupper av elever, i förhållande till elevers kön och etnicitet, menar ungefär en eller två lärare av tio. Att proven diskriminerar elever med annat modersmål än svenska är en vanligare uppfattning bland lärare som undervisar i denna typ av klasser.

Som elev är man osäker på vad proven egentligen betyder i relation till betyget. Man pekar på att det skiljer både mellan ämnen och mellan lärare hur resultat på nationella prov värderas. En allmänt uttryckt känsla bland eleverna är att det spelar större roll hur det går på de nationella proven än på de lokala proven. Att proven just är nationella, dvs. att alla elever i hela landet gör samma prov, är också något som lyfts fram generellt av eleverna som en positiv kvalitet. För öv-

Dnr 01-2003:2038

rigt varierar elevernas uppfattningar om själva proven, som t.ex. med avseende på ämne, delprov, innehåll, kravnivåer eller form.

Flertalet rektorer och lärare beskriver att provresultaten används för pedagogisk uppföljning (uppges oftare av lärare i år 5) och för diskussioner med olika intressenter, oftast med elever och kollegor.

Flertalet lärare uppger att man har goda möjligheter till samarbete och stöd bland sina kollegor om provens användning. Däremot är det inte alls vanligt med samverkan mellan skolor om proven.

Utredningsgruppens bedömning

Utredningsgruppen bedömer att rektorer och lärare i grundskolan är nöjda med det existerande provsystemet och tycker att det väl fyller sina syften. Också eleverna förefaller nöjda även om de ger uttryck för viss osäkerhet om provens betydelse för deras betyg. Lärare och rektorer är inte intresserade av att ändra något i systemets struktur som till exempel att utvidga det att omfatta fler ämnen och åldersgrupper.

3. Några jämförelser av provbetyg och slutbetyg

Ett viktigt syfte med de nationella proven är att stödja en likvärdig bedömning och rättvis betygssättning. Vad detta innebär för relationen mellan provbetyg och slutbetyg²⁶ är inte helt enkelt att bedöma. I skolans styrdokument uttrycks dock klart att de nationella proven inte kan pröva alla mål och att det är läraren som har ansvaret för att betygssätta eleverna. Därvid ska hänsyn tas till alla elevens prestationer och de nationella proven ska inte tillmätas särskild vikt vid betygssättningen av den enskilda eleven. En enskilds elevs slutbetyg kan alltså avvika såväl uppåt som neråt i relation till provbetyget. Det beror inte minst på att prov är osäkra som mätinstrument, dvs. de har en betydande felmarginal (se appendix för en närmare beskrivning).

För en enskild elev är det således inte anmärkningsvärt om provbetyg och slutbetyg skiljer sig åt. När det däremot gäller gruppjämförelser kan eventuella systematiska skillnader inte på samma sätt förklaras med provens mätfel. Om skillnader förekommer i sådana fall måste de, om de ligger utanför vad som kan betraktas som slumpmässigt, kunna förklaras på annat sätt.

I det här avsnittet redovisas några jämförelser mellan provbetyg och slutbetyg för grundskolans ämnesprov och för de obligatoriska kursproven i gymnasieskolans kärnämnen. Jämförelserna bygger på de slumpmässiga urval av skolor som Skolverket årligen har som underlag för att redovisa provresultat.²⁷ För ämnena svenska och svenska som andraspråk och engelska har provresultat valts från år 2001. Detta beror på ett beslut av Skolverket 2001, att från och med proven år 2002 ska inte några sammanfattande provbetyg redovisas i dessa ämnen utan endast betyg på de olika delproven (tala, läsa och skriva). I matematik redovisas dock fortfarande provbetyg och för det ämnet används därför de senaste resultaten som finns tillgängliga, nämligen från vårterminen 2002.

De jämförelser som redovisas innebär dels en beskrivning av skillnaden mellan slutbetyg och provbetyg uppdelat på de skolor som ingår i urvalet, dels en beskrivning av betygsfördelningen på provet och i slutbetyget för hela gruppen. Endast de elever som har både provbetyg och slutbetyg ingår i jämförelserna. Skolor med färre än 20 redovisade elevresultat ingår inte heller i underlaget.

²⁶ ”Slutbetyg” ska i det här sammanhanget förstås som det betyg eleven får i respektive ämne i sitt slutbetyg från grundskolan, eller i respektive kurs för gymnasieskolans del.

²⁷ Gäller i cirka 150 skolor och 10 000 elever i grundskolan samt ungefär 100 skolor och cirka 10 000 – 15 000 elever beroende på ämne. Resultat insamlas också från cirka 50 komvuxenheter men de resultaten ingår inte i den här redovisningen. För en utförligare resultatredovisning hänvisas till Skolverket årliga publikationer av provresultat.

Dnr 01-2003:2038

Det är också viktigt att framhålla att bortfallet i resultatredovisningarna är betydande. För grundskolans del saknas resultat för cirka 15 procent av eleverna i engelska och cirka 17-18 procent av eleverna i både svenska/ svenska som andraspråk och i matematik. Till detta ska läggas att av de skolor och lärare som rapporterat provresultat är det en rätt stor andel som inte gjort alla provdelar. Av redovisade resultat har 17, 18 och 19 procent i respektive engelska, matematik och svenska inte kompletta provbetyg. Det betyder att fullständiga provresultat endast föreligger för drygt två tredjedelar av eleverna i urvalet.

För gymnasieskolans del är bortfallet ännu större. Inrapporterade resultat saknas för knappt 30 procent av eleverna i svenska B, för cirka 25 procent av eleverna i matematik A och för knappt 20 procent av eleverna i engelska A. Den andel av de rapporterade eleverna som inte har fullständiga provresultat skiljer sig betydligt mer mellan ämnena än vad fallet är i grundskolan. I svenska är det 45 procent inrapporterade eleverna som inte har fullständiga provbetyg, i engelska 16 procent och i matematik 9.²⁸ De aktuella proven är obligatoriska sedan 1 juli 2001 och det är uppenbart att många skolor och lärare bryter mot gällande föreskrifter.

När det gäller slutbetyg är grundskoleelevernas betyg hämtade från SCBs insamling medan slutbetygen (kursbetygen) för gymnasieeleverna är de *preliminära* betyg lärarna inrapporterat. I det senare fallet kan således avvikelser förekomma mellan preliminära och slutgiltiga betyg. Anledningen till att preliminärbetyg används är att SCB inte har tillgång till kursbetyg förrän efter det att eleven avslutat gymnasieskolan, dvs. först ett par, tre år efter provgenomförandet.

I samtliga jämförelser ingår endast de elever som har *både* provbetyg och slutbetyg (ämnesbetyg eller kursbetyg).

Syftet med det här kapitlet är främst att undersöka om relationen mellan provresultat och slutbetyg skiljer sig åt för olika skolor (och för olika program när det gäller gymnasieskolan) och mellan olika ämnen. Försiktighet i tolkningarna bör naturligtvis anbefallas med tanke på de stora bortfallet av rapporterade resultat, i synnerhet för gymnasieskolan.

Grundskolan

För enskilda elever kan som nämnts provresultat variera en hel del, dels mellan olika provtillfällen, dels i relation till vad eleven "egentligen" kan. Det beror som nämnts på att prov är osäkra som mätinstrument. Det finns därför inte skäl att förvänta sig att resultaten på

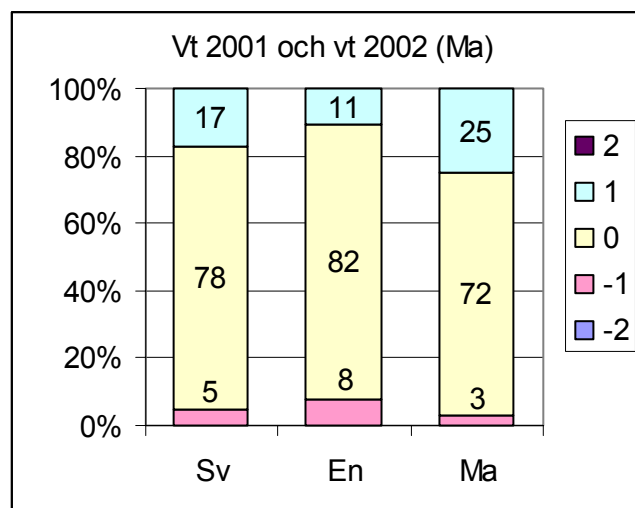
²⁸ En möjlig förklaring till skillnaderna mellan grundskolan och gymnasieskolan kunde vara att grundskolans delprov genomförs vid olika tillfällen, medan gymnasieskolans prov genomförs vid ett och samma tillfälle samt att gymnasieskolan anser sig ha snävare tidsramar och därmed mindre tid för provverksamhet.

Dnr 01-2003:2038

två olika prov, t.ex. i form av provbetyg, ska sammanfalla för *alla* elever. För flertalet kommer de att sammanfalla men somliga kommer att lyckas relativt sett bättre på det ena provet och andra på det andra provet. Ju reliablare proven är desto större är överensstämmelsen, men slumpfaktorer kommer alltid att medföra variationer. Det samma gäller förhållandet mellan provbetyg och slutbetyg. Även om man skulle förutsätta att de mäter samma kunskaper - vilket de endast gör till viss del - skulle betygen inte överensstämma för somliga elever. *Hur* stora avvikelserna kan vara innan det finns anledning till oro för provens mätsäkerhet är däremot svårt att avgöra. Likaså kan det vara svårt att avgöra hur *små* de bör vara innan man kan misstänka att proven fått avgöra slutbetyget i alltför stor utsträckning.

Jämförelser mellan ämnen

Utan att värdera utfallet kan man konstatera att överensstämmelsen mellan provbetyg och slutbetyg ser ut som i figur 23 för de elever som gjort ämnesproven i år 9 och som ingår i urvalet. Figuren anger alltså hur stor andel av eleverna som de facto har samma respektive olika provbetyg och slutbetyg. Markeringen "1" anger att slutbetyget är ett steg högre än provbetyget, "0" att det är lika och "-1" att slutbetyget är ett steg lägre.²⁹ Så har t.ex. 17 procent av eleverna i svenska högre slutbetyg än provbetyg, 78 procent har samma betyg och 5 procent har lägre slutbetyg.



Figur 23. Andel av eleverna i årskurs 9 som erhållit samma prov- och slutbetyg, eller betyg som skiljer sig åt ett (eller två) steg.

Vad man kan konstatera är att ungefär tre fjärdedelar av eleverna har samma slutbetyg som provbetyg. En höjning av betyget är vanligast i matematik och minst vanligt i engelska. Engelska är också det ämne där höjningar och sänkningar av betyg är mest lika, med viss över-

²⁹ Avvikelser med två betygssteg är så ovanliga, cirka en elev av tusen, att de inte ger utslag i figuren.

Dnr 01-2003:2038

vikt för höjning, medan matematik är det ämne där andelen höjda betyg är störst och andelen sänkta minst.

Jämförelser mellan skolor

Den vänstra av de parvisa figurer som följer torde inte kräva någon närmare förklaring. De olika delarna av respektive stapel i t.ex. figur 24 anger hur många procent av eleverna som har de olika betygen.

I den högra figuren representerar varje stapel en skola. Figuren avser att visa hur många procent av eleverna som har slutbetyg som avviker från provbetyg med ett steg. Stapeln längst till vänster i figur 25 anger t.ex. att vid den skolan har drygt 40 procent av eleverna ett slutbetyg som ligger ett steg högre än provbetyget. Stapeln längst till höger anger på motsvarande sätt att vid den skolan har cirka 20 procent av eleverna slutbetyg som ligger ett steg lägre än provbetyget.

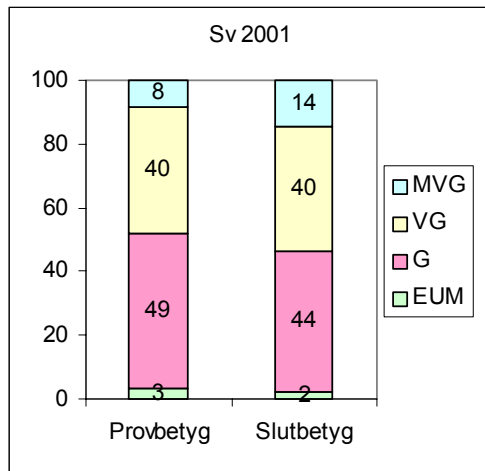
Observera att för de diagram som följer handlar det om genomsnittsvärden (till skillnad mot vad som gäller i figur 1 där både höjning och sänkning anges). Det betyder att om t.ex. 50 procent av eleverna har ett slutbetyg som ligger ett steg högre än provbetyget och 10 procent ett slutbetyg som ligger ett steg lägre så blir nettoresultatet, eller genomsnittet, för skolan att 40 procent av eleverna har ett slutbetyg som ligger ett steg högre än provbetyget (dvs. vad som motsvarar skillnaden mellan den andel som höjt respektive sänkt betyg i föregående figur).

Resultaten av jämförelserna redovisas först för de tre provämnena och kommenteras sedan i ett avslutande avsnitt.

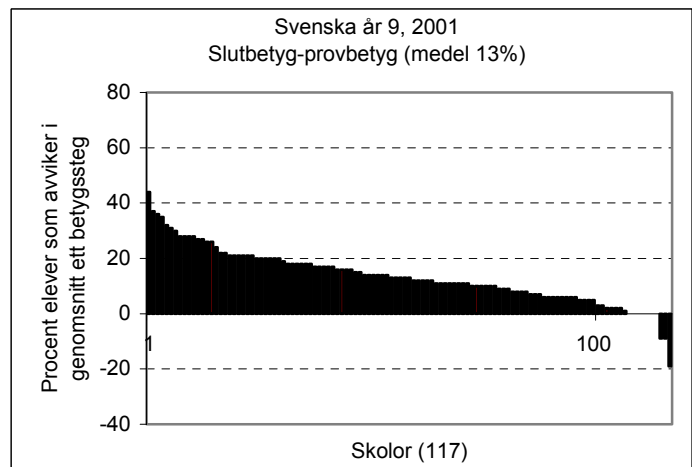
Svenska

I redovisningen ingår endast elever som har betyg i svenska. Elever som läser svenska som andraspråk genomför samma prov men är inte medräknade i betygsredovisningarna (såvida de inte också har betyg i svenska).

Dnr 01-2003:2038



Figur 24. Fördelning av provbetyg och slutbetyg i svenska år 9, vt 2001.



Figur 25. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av slutbetyg och provbetyg, uppdelat på skolor.

I figur 24 kan man konstatera att framför allt andelen elever med MVG är större när det gäller slutbetyget än provbetyget. Andelen som har G är å andra sidan ungefär motsvarande andel lägre. Cirka sex procent (netto) av eleverna "höjer"³⁰ sitt betyg från VG till MVG och att ungefär lika stor andel höjer från G till VG. En procent höjer från att inte ha nått målen till att få G. Sammantaget innebär det netto är cirka 13 procent av eleverna höjer sitt betyg ett steg.

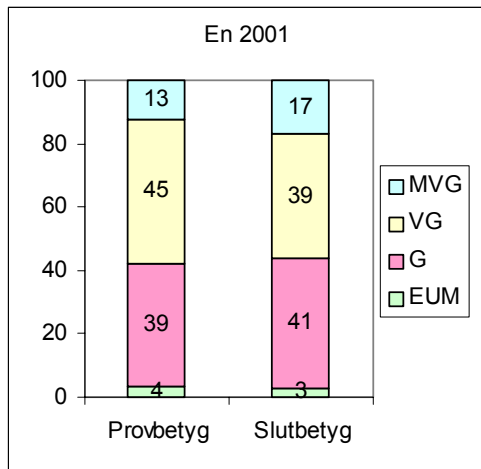
Figur 25 visar samma sak ur ett annat perspektiv. Den visar att vid flertalet skolor får elever högre slutbetyg än provbetyg. Den genomsnittliga nettoeffekten är även här 13 procent (dock beräknad som medelvärde för de ingående skolorna utan viktning för olika elevantal). Figuren visar därtill att variationen mellan skolor kan vara rätt stor. Från de skolor längst till vänster i figuren där upp emot 40 procent av eleverna har slutbetyg som ligger ett steg högre än provbetyget, till den skola längst till höger där ungefär 20 procent av eleverna har *lägre* slutbetyg än provbetyg.

Engelska

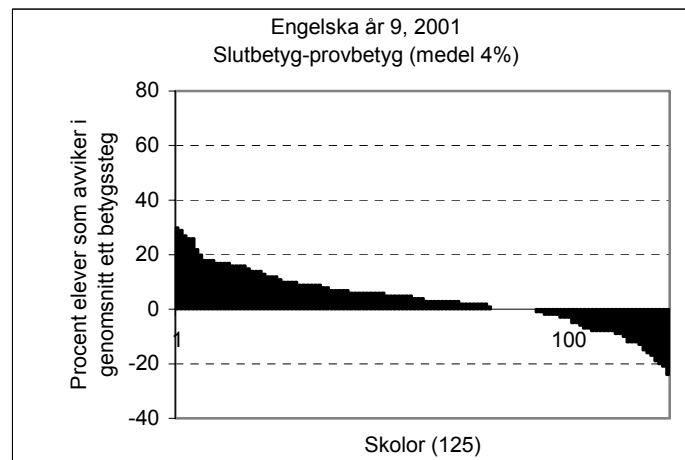
Även för engelskans de gäller resultaten vårterminen 2001.

³⁰ Det är naturligtvis tveksamt att tala om att "höja" betyget eftersom de två betygen inte gäller samma saker. Slutbetyget innefattar även mål som inte ingår i provet. Även här gäller det nettoeffekten, det behöver alltså inte vara så att sex procent höjt betyget och noll procent sänkt det, utan det kan t.ex. vara så att 10 procent höjt sitt betyg och fyra procent sänkt det. Nettoeffekten blir sex procents höjning.

Dnr 01-2003:2038



Figur 26. Fördelning av provbetyg och slutbetyg i engelska år 9, vt 2001.



Figur 27. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av slutbetyg och provbetyg.

Man kan konstatera att mönstret i engelska är delvis annorlunda än i svenska. Figur 26 visar att visserligen tycks cirka fyra procent av eleverna gå från VG till MVG, men samtidigt minskar andelen elever med VG sex procentenheter vilket innebär att två procent av eleverna sänker sitt betyg till G, samtidigt som en procent höjer från att inte nå målen på provet till G i slutbetyg. Det ger en nettohöjning på cirka tre procentenheter.

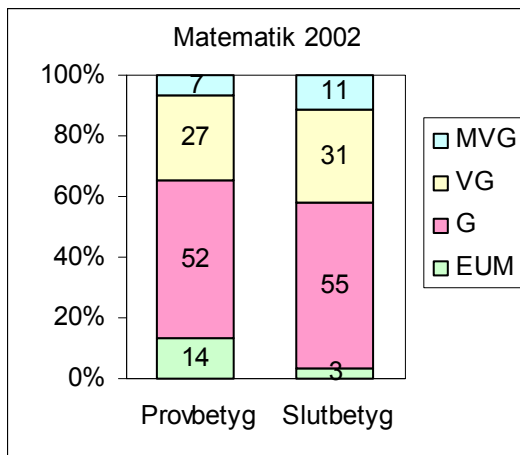
Figur 27 visar att bilden är mer varierad i engelska än i svenska. En svag nettohöjning på fyra procent kan noteras³¹ men för ungefär var fjärde skola gäller att slutbetygen ligger lägre än provbetygen. Skillnaden mellan de skolor som ligger längst till vänster i figuren och de som ligger längst till höger uppgår dock till cirka 50 procentenheter och man kan naturligtvis undra över vad som döljer sig bakom dessa siffror. Varför får var tredje elev högre slutbetyg än provbetyg vid en skola medan nettoeffekten vid en annan skola blir att var tredje elev får lägre slutbetyg än provbetyg? Är detta förenligt med provens syfte att stödja en rättvis och likvärdig betygssättning?

Matematik

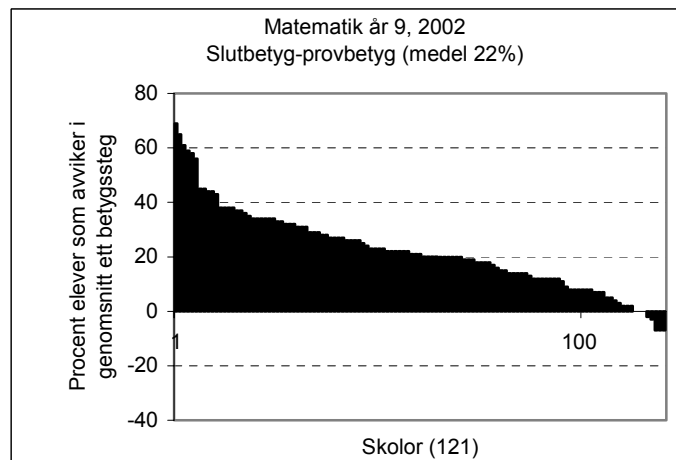
I matematik kvarstår provbetyg fortfarande och därför används i redovisningen det senaste tillgängliga provresultatet från vårterminen 2002. Anledningen till att provbetyget är kvar i matematik är att det där inte ansågs möjligt att dela upp provet i avgränsade delar på det sätt som kunde göras i svenska och engelska.

³¹ Skolorna är ej viktade efter antal elever, vilket kan ge vissa mindre avvikelser mellan andelarna som har högre slutbetyg (tre respektive fyra procent för engelska).

Dnr 01-2003:2038



Figur 28. Fördelning av provbetyg och slutbetyg i matematik år 9, vt 2002.



Figur 29. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av slutbetyg och provbetyg.

Mönstret i matematik (figur 28, 29) är mer likt mönstret i svenska än i engelska. Andelen elever som ökar från VG till MVG är dock densamma som i engelska (4 procent) och lägre än i svenska (6 procent). Andelen som ökar från G till VG är dock dubbelt så stor, nämligen 8 procentenheter, och andelen elever som går från att inte ha nått målen till att bli godkända är 11 procentenheter. Nettoandelen elever som har ett slutbetyg som ligger ett betygssteg högre än provbetyget är således cirka 23 procent (22 procent sett som skolgenomsnitt).

Skillnaderna mellan provbetyg och slutbetyg är störst i matematik och minst i engelska med svenska i ett mellanläge. Man kan också notera att andelen elever med slutbetygen VG eller MVG är betydligt lägre i matematik (42 procent) än i engelska (56 procent) och svenska (54 procent).

Är då de uppvisade mönstren specifika för de aktuella proven? En genomgång av samtliga genomförda prov sedan de startade 1998 visar att mönstret med vissa smärre variationer är mycket likartat för de prov som genomförts hittills.

Några frågor som naturligt infinner sig är dels hur de skillnader som kan konstateras mellan provbetyg och slutbetyg kan förstås, dels hur skillnaderna mellan olika ämnen kan förklaras. Några försök till förklaringar kommer att ges i det avslutande avsnitt, men först redovisas några resultat från gymnasieskolan.

Gymnasieskolan

En jämförelse av provbetyg och kursbetyg (slutbetyg i kursen) är mer komplicerad för gymnasieskolan än för grundskolan. Elevsammansättning och programsammansättning kan variera i betydande grad mellan olika gymnasieskolor och det är därför tveksamt i vilken utsträckning en jämförelse rakt av mellan skolor ger en rimlig och rättvisande bild.

Dnr 01-2003:2038

Samtidigt är dock kursmål och betygskriterier desamma för en viss kurs oberoende av inom vilket program den genomförs. Det finns därför ur det perspektivet ingen anledning att anta att relationen mellan provbetyg och slutbetyg skulle skilja sig åt mellan olika program (och olika skolor). Men detta är en teoretisk utgångspunkt och det kan finnas skäl att undersöka hur det förhåller sig. Redovisningen av resultat för gymnasieskolan kommer därför inte endast att delas upp på olika skolor utan också efter olika program. Dock inte samtidigt utan i ena fallet slås alla program vid en skola samman till ett "skolvärde". I det andra fallet slås alla skolor samman vid konstruktionen av "programvärden".

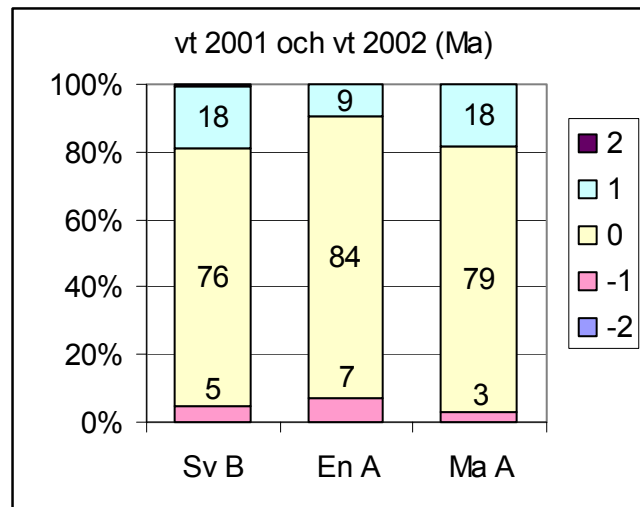
Framställningen kan givetvis ifrågasättas på olika sätt, inte minst med tanke på det förhållandevis stora bortfallet av resultat, men den kan trots detta tjäna som underlag för en diskussion. Intresset i den här framställningen är fokuserat på relationen provbetyg – kursbetyg och betygsskillnader mellan olika program i övrigt, t.ex. andel elever i respektive program med betyget IG, diskuteras inte närmare, även om sådana resultat givetvis har ett intresse. Ytterligare information om kursproven finns som tidigare nämnts i Skolverkets årliga rapporter om resultaten, t.ex. "Gymnasieskolans kursprov vårterminen 2001" (Skolverket 2001).

Det bör också påminnas om att slutbetyg i det här sammanhanget avser de preliminära kursbetyg lärarna uppgett vid resultatredovisningen.

Jämförelse mellan ämnen

Redogörelsen för gymnasieskolan gäller samma ämnen som i grundskolan, dvs. svenska/svenska som andraspråk (sammanslaget, varav 2 procent har betyg i svenska som andraspråk), engelska och matematik. För kärnämneskurserna svenska B, engelska A och Matematik A visar figur 30 hur stor andel av eleverna som de facto har samma kursbetyg och provbetyg respektive kursbetyg som är ett betygssteg högre eller lägre än provbetyget.

Dnr 01-2003:2038



Figur 30. Andel av eleverna i gymnasieskolan som erhållit samma prov- och kursbetyg, eller betyg som skiljer sig åt ett (eller två) steg.

Man kan konstatera att mönstret i stor utsträckning sammanfaller med det som gällde för grundskolan (fig. 23), framför allt i svenska och engelska. För matematikens del är avvikelserna något mindre för gymnasieskolan, främst genom att andelen med högre slutbetyg än provbetyg är mindre än för grundskolan.

Jämförelser mellan skolor och program

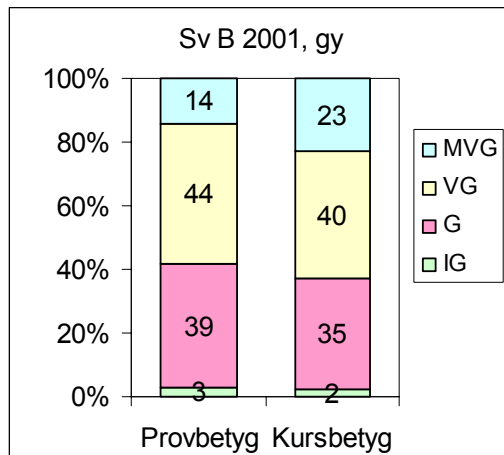
Gymnasieskolan är mer heterogen och uppdelad än grundskolan och det kan därför vara av intresse att undersöka inte bara skillnaden mellan olika skolor utan också mellan olika program.

Svenska

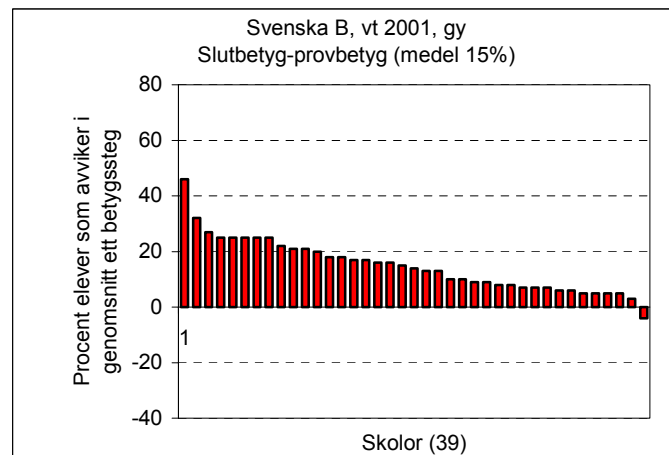
I svenska finns det endast ett nationellt prov på gymnasienivå, nämligen i kärnämneskursen svenska B. Även för gymnasieskolan gäller att inget sammanfattande provbetyg ges från och med 2002 och därför används provet med sammanfattande provbetyg från vårterminen 2001.

Sammanvägda provbetyg finns för en förhållandevis liten andel elever i svenska kurs B. Detta beror på att endast 55 procent av eleverna gjort samtliga delprov. Detta beror i sin tur på att många lärare väljer att inte göra delprov A (ett mer styrt prov som leder till en kort, koncentrerad text) och de muntliga delprovet C. Flertalet (85 procent) genomför däremot delprov B som är en mer traditionell skrivuppgift ("uppsats").

Dnr 01-2003:2038



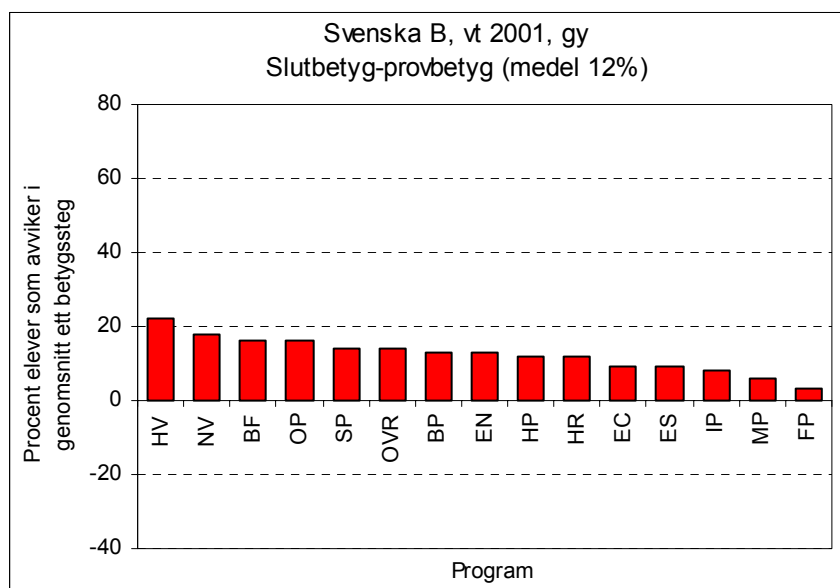
Figur 31. Fördelning av provbetyg och kursbetyg i svenska B, vt 2001.



Figur 32. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av kursbetyg och provbetyg, svenska B, vt 2001.

Mönstret från grundskolans nationella prov känns igen gymnasieskolans betyg i svenska kurs B (figur 31, 32). Det är dock en större andel elever som i gymnasieskolan har högre kursbetyg än provbetyg. En knapp fjärdedel av eleverna har MVG i slutbetyg vilket är 9 procentenheter mer än för provbetyget. I genomsnitt höjs betyget ett steg för 15 procent av eleverna (och skolorna), medan variationen mellan skolor går från höjning för cirka var tredje elev (nästan hälften om man ser på skolan längst till vänster i figuren) ner till i stor sett samma fördelning för båda betygen.

När det gäller skillnaden mellan kursbetyg och provbetyg uppdelat på program blir mönstret som i figur 33.



Figur 33. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av kursbetyg och provbetyg i svenska B, vt 2001, uppdelat på program.

Dnr 01-2003:2038

Figuren visar skillnaden mellan slutbetyg (kursbetyg) för de olika programmen.³² Differensen i genomsnittlig skillnad (12 procent) gentemot skolkjämförelserna (15 procent) kan hänföras till att medelvärdet för program inte är viktat. De i elevantal klart dominerande programmen SP och NV ligger på den halva av program där skillnaden mellan kursbetyg och provbetyg är större än medelvärdet 12 (se tabell 8 för antal elever).

Att vissa program har så få grupper (och elever) i urvalet gör att resultaten i figur 33 får tolkas med försiktighet. Livsmedelsprogrammet ingår inte i figuren eftersom det aktuella urvalet endast innefattar 2 grupper med sammanlagt 9 elever.

Trots att vissa program har få grupper tyder mönstret i figur 33 på att relationen kursbetyg – provbetyg är tämligen likartad för olika program. Skillnaderna mellan olika skolor är större än skillnaden mellan program (med reservation för att en sådan jämförelse inte är helt rättvisande).

Det bör också påpekas att den bild som ges av att relationen mellan kursbetyg och slutbetyg är

Tabell 8. Antal elever och grupper som ingår i urvalet av resultat Sv B vt 2001.

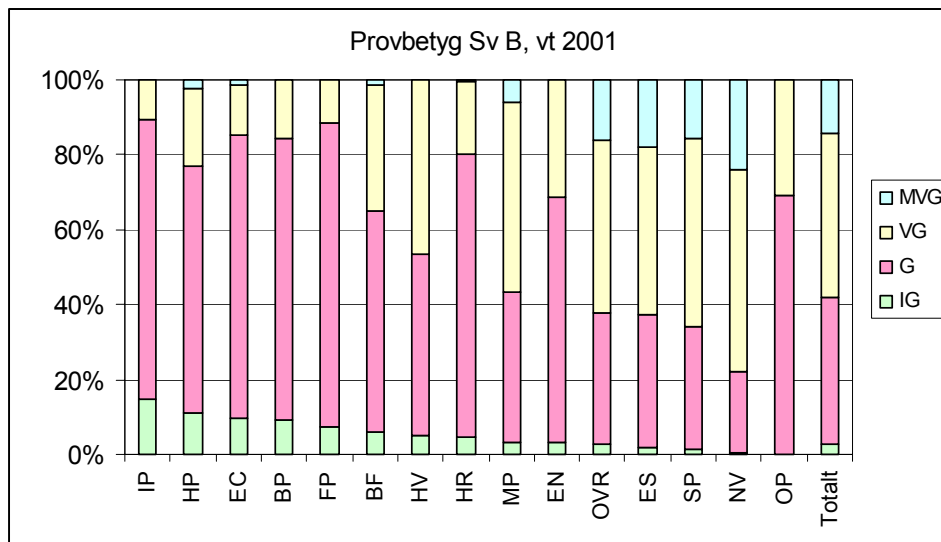
Program	Elever	Grupper
BF	246	15
BP	107	12
EC	199	15
EN	32	6
ES	210	10
FP	69	11
HP	82	9
HV	60	8
HR	147	10
IP	95	13
LP	9	2
MP	155	8
NV	1227	28
OP	81	7
SP	1568	31
OVR	484	22
Totalt	4771	54

förhållandevis likartad mellan program inte ska uppfattas som att betygsfördelningarna för olika program är likartade. I det fallet är skillnaderna stora.

Figur 34 visar till exempel att knappast några elever på yrkesinriktade program får provbetyget MVG, med undantag för estetiska programmet och medieprogrammet. Det är elever på dessa program samt på de naturvetenskapliga och samhällsvetenskapliga programmen som uppnår detta betyg. (Kategorin övriga, "OVR", utgörs främst av olika specialutformade program, som ofta kan ses som varianter av de studieförberedande programmen).

³² "Övriga" innefattar program med få elever, främst olika specialutformade program.

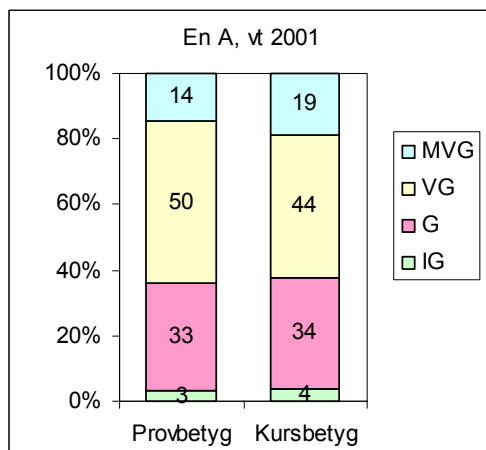
Dnr 01-2003:2038



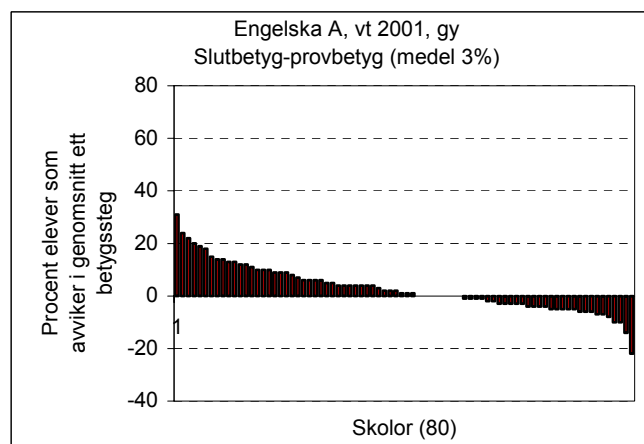
Figur 34. Fördelning av provbetyg i svenska B, vt 2001, uppdelat efter program.

Engelska

För engelskans del redovisas resultaten för provet i kärnämneskursen engelska A. Av de redovisade resultaten (15 762 elever) har 84 procent gjort samtliga delprov och fått ett provbetyg. 82 procent har både provbetyg och kursbetyg varför underlaget är alltså betydligt större i engelska, och därmed mer tillförlitligt, än i svenska. Fördelningen av provbetyg och kursbetyg visas i figur 35



Figur 35. Fördelning av provbetyg och kursbetyg i engelska A, vt 2001

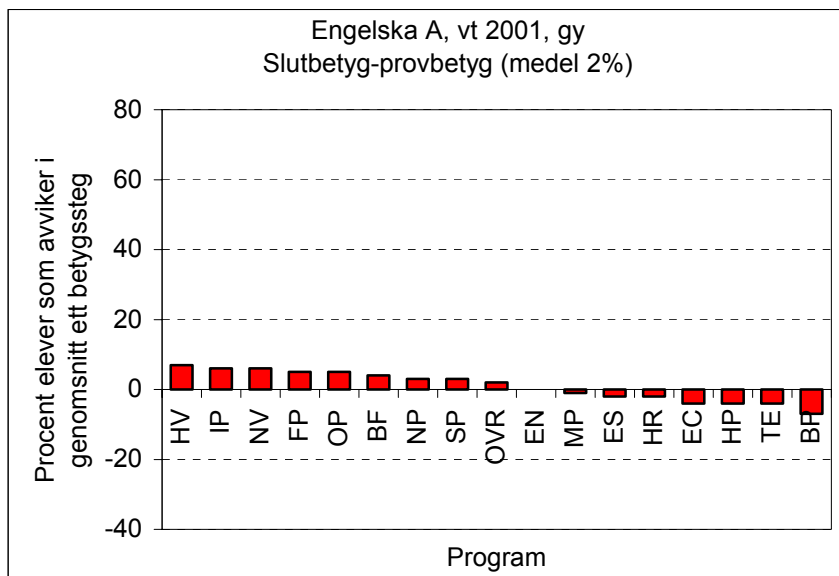


Figur 36. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av kursbetyg och provbetyg, engelska A, vt 2001.

Även för engelskans del kan man vid en jämförelse med grundskolans resultat konstatera att mönstret är påfallande likartat. Den andel av eleverna som har MVG ökar medan andelen med VG minskar. Andelen elever som inte nått målen i grundskolan respektive fått IG i gymnasieskolan är ungefär lika stor och relationen mellan provbetyg och slutbetyg är i stort sett densamma för båda skolformerna.

Dnr 01-2003:2038

Även på skolnivå är bilden mycket likartad (figur 36). Vid flertalet skolor är kursbetygen något högre än provbetygen, men antalet skolor där förhållandet är det omvända är nästan lika många.



Figur 37. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av kursbetyg och provbetyg i engelska A, vt 2001, uppdelat på program.

Uppdelat på program (figur 37) blir bilden likartad den som gäller på skolnivå med undantag av att skillnaden mellan program är mindre än mellan skolor.

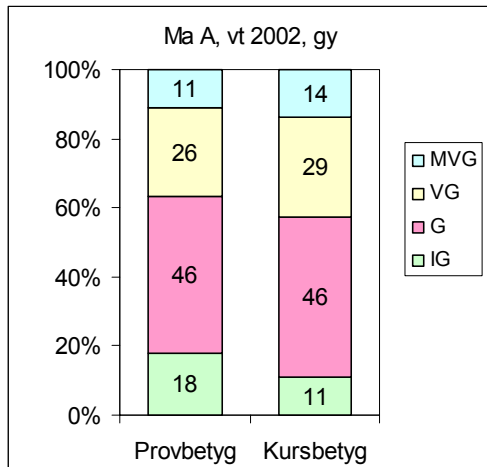
Matematik

För matematikens del finns nationella prov för kärnämneskursen matematik A, men även för karaktärsämneskurserna B, C och D. Därtill finns prov för matematik E i provbanken.

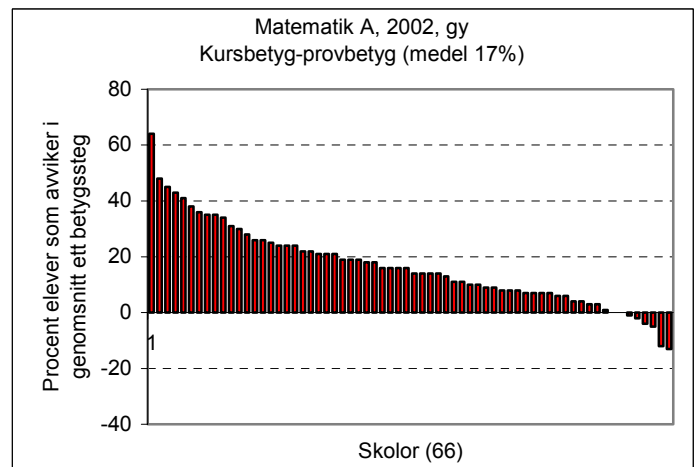
I den här redovisningen ingår de senaste tillgängliga resultaten från matematik A vårterminen 2002. Av det totalt inskickade resultaten för 9826 elever föreligger både provbetyg och kursbetyg för 8768 elever eller 89 procent.

Figur 38 visar fördelningen av provbetyg och kursbetyg. Mönstret påminner även för matematik om det som gällde för grundskolan, men en betydligt större andel av gymnasieeleverna når inte målen och andelen elever som har betyget IG både på provet respektive som slutbetyg är större.

Dnr 01-2003:2038



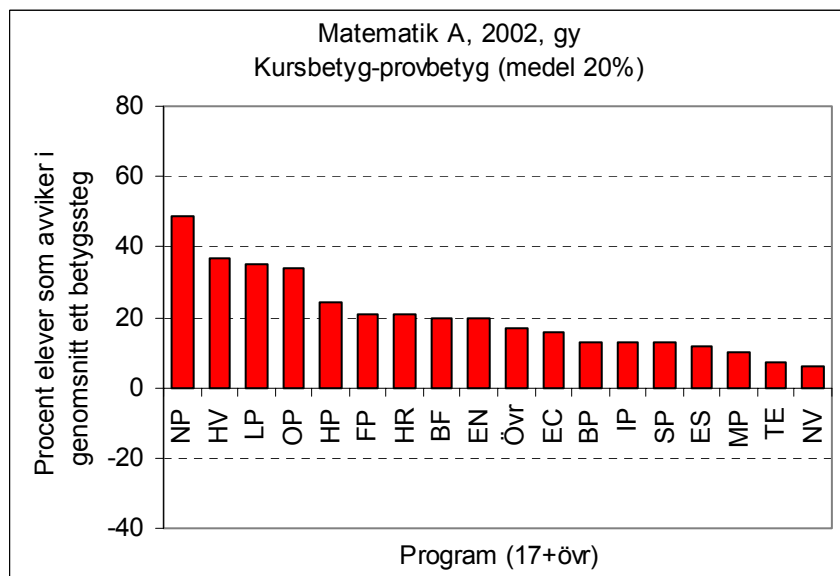
Figur 38. Fördelning av provbetyg och kursbetyg i matematik A, vt 2002



Figur 39. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av kursbetyg och provbetyg, matematik A, vt 2002.

Liksom för grundskolan är skillnaden mellan slutbetyg och provbetyg störst för matematikämnet (figur 39). Även variationen mellan skolor är stor. För vissa skolor är kursbetyget ett steg högre än provbetyget för nästan varannan elev, men det finns enstaka skolor där kursbetygen är lägre än provbetygen.

Även när det gäller de olika programmen är skillnaderna större för matematik än för de andra provämnena (figur 40).



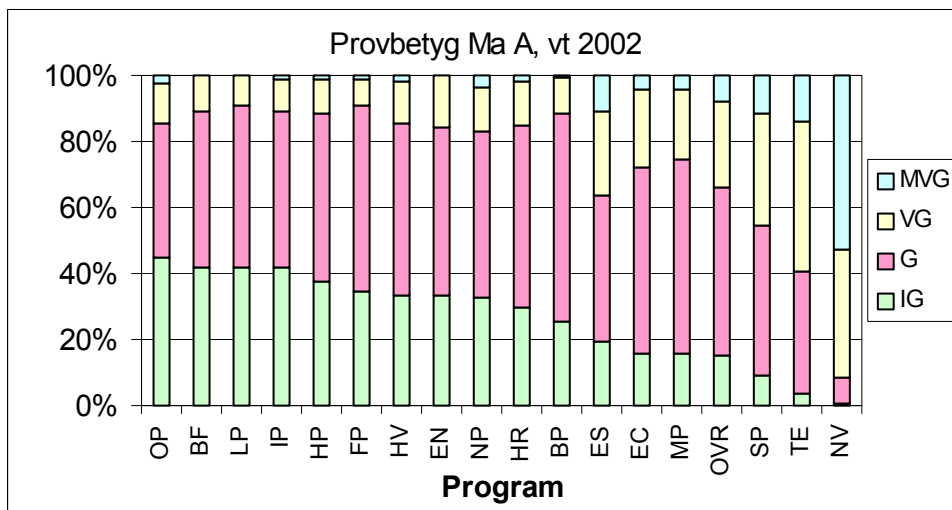
Figur 40. Andel elever som i genomsnitt avviker ett betygssteg vid jämförelse av kursbetyg och provbetyg i matematik A, vt 2002, uppdelat på program.

Man kan här notera att skillnaderna är störst för de yrkesinriktade ämnena och minst för de studieförberedande, om man till dessa räknar det estetiska programmet och medieprogrammet. Matematik skiljer sig i det här avseendet från svenska och engelska där någon sys-

Dnr 01-2003:2038

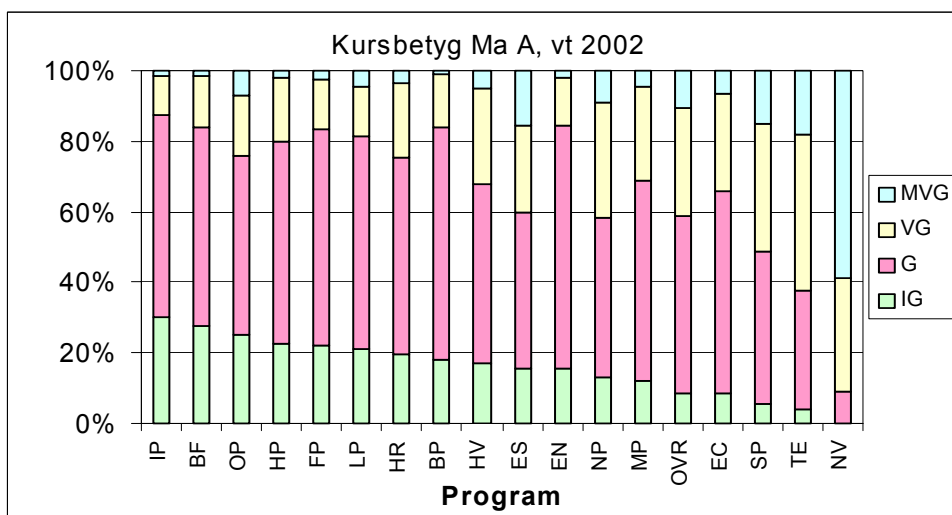
tematisk skillnad mellan yrkesinriktade och studieförberedande program inte kan noteras.

Skillnaden för de yrkesinriktade programmen beror främst på att det är andelen som har IG på provet som är större än andelen som har IG i kursbetyg. Figur 41 visar fördelningen av provbetygen för de olika programmen.



Figur 41. Fördelning av provbetyg i matematik A, vt 2002, uppdelat efter program.

Figur 42 visar motsvarande fördelning av (de preliminära) kursbetygen för samma elever.



Figur 42. Fördelning av kursbetyg i matematik A, vt 2002, uppdelat efter program.

Diskussion

Sammanfattningsvis kan man konstatera att det finns ett antal olikheter som kan vara intressanta att diskutera. Det gäller

Dnr 01-2003:2038

- a) skillnader mellan olika ämnen,
- b) skillnader mellan olika skolor,
- c) skillnader mellan olika program,
- d) skillnader mellan grundskola och gymnasieskola.

För att få ett underlag för denna diskussion är det dock först nödvändigt att föra en mer allmän diskussion om relationen mellan provbetyg och slutbetyg. En sådan diskussion inleder därför avsnittet. Därefter diskuteras de fyra punkterna ovan varefter avsnittet avslutas med ett försök att bedöma vilka konsekvenser - om några - de här redovisade resultaten kan ha för utformningen av ett framtida provprogram.

Är provbetyg och slutbetyg jämförbara?

Vad representerar provbetyg och vad representerar slutbetyg? Är de mått på samma saker eller är de mått på olika saker? Vad ligger till grund för respektive betyg? Det är några i sammanhanget väsentliga frågor.

Provbetygen

De nationella proven konstrueras som tidigare beskrivits vid olika universitetsinstitutioner i samverkan med praktiskt verksamma lärare. Uppgifter konstrueras, utprövas, analyseras och diskuteras noggrant innan de slutgiltiga proven är färdiga. Vissa elevlösningar från utprövningarna väljs ut och diskuteras utförligt i relation till kursmål och betygskriterier för att utgöra exempel på lösningar av olika kvalitet, dvs. lösningar som representerar olika betyg.

Om ett prov (eller delprov) som består av uppgifter ska resultera i ett provbetyg måste de olika uppgiftsresultaten vägas samman på något sätt. Det kan ske på olika sätt. Det vanligaste i provsammanhang är att man poängsätter uppgifterna så att man får en poängskala för provet. Därefter fastlägger man vissa poänggränser för de olika betygen. Eventuellt kan man föreskriva att t.ex. en viss andel poäng måste inhämtas på VG-nivå för att ett sådant provbetyg ska uppnås (det gäller t.ex. i matematik):

En sådan gränssättning på en poängskala (standard setting) är egentligen inte i enlighet med det målrelaterade system vi har. Provpoängen anger inte vilka mål som uppnåtts och vilka som eventuellt inte uppnåtts. Provpoängen kan sägas vara kompensatorisk, god målfyllelse för vissa mål kan kompensera otillräcklig målfyllelse i andra avseenden. Denna metod är dock mer eller mindre nödvändig av praktiska skäl. Om uppgifter skulle konstrueras för att pröva varje delmål med betryggande mätsäkerhet skulle proven bli så omfattande och tidskrävande att de inte skulle vara möjliga att genomföra. Låt oss alltså tills vidare acceptera att proven inte kan pröva alla specifi-

Dnr 01-2003:2038

ka delmål utan endast ge en mer övergripande bild av måluppfyllelse, t.ex. inom målområden som läsförståelse, skrivkunnande, muntlig förmåga etc.

Hur fastställs då poänggränserna för de olika betygen på de nationella proven (delproven)? Vanligen görs de med någon form av ”modifierad Angoff” i kombination med mer holistiska bedömning³³, dvs. en bedömning av vad som utifrån en tolkning av mål och kriterier och utifrån den egna erfarenheten bedöms som rimliga gränser för olika betyg. Till sitt stöd kan bedömaregruppen ibland också utnyttja utvärderingsdata. Bedömningen görs av en särskilt rekryterad och instruerad grupp av cirka 10 till 15 lärare. Till sin hjälp kan dessa också ha visst underlag från utvärderingar.

Provbetygen kan alltså sägas vara resultaten av dessa 15 lärares samlade och samordnade bedömning av vad elever bör prestera på provet för att erhålla respektive betyg.

Slutbetyget i ämnet eller kursen

Slutbetyget sätts av läraren på basis av det samlade underlag läraren har om elevens prestationer. För att eleven ska bli godkänd ska enligt grundskoleförordningen (7 kap, 8§) i princip alla mål vara uppnådda för att en elev ska bli godkänd. Endast om särskilda skäl av typen ”funktionshinder eller andra liknande personliga förhållanden som inte är av tillfällig natur” föreligger får undantag göras. Det innebär (implicit) att för övriga elever ska alla mål vara uppnådda för att eleven ska bli godkänd. Elevernas ska betygssättas i enlighet med de mål och betygskriterier som gäller för kursen. Beroende på stoffval kan (bör) dessa ha olika lokal utformning men de måste på ett tydligt sätt kunna kopplas till de nationella målen och betygskriterierna. Det är alltså i princip lärarens samlade bedömning av elevens måluppfyllelse för varje enskilt mål som avgör elevens betyg. Om ett av målen inte uppnåtts ska eleven enligt föreskriften inte bli godkänd. Det är en kategorisk och sträng regel.

Relationen provbetyg - slutbetyg

Vad brukar då sägas om slutbetyget och dess relation till provbetyget? Någon särskilt omfattande debatt har inte förts, men när frågan om varför slutbetyg och provbetyg inte överensstämmer förs på tal brukar ett antal alternativa förklaringar framföras. I allmänhet går det ut på att förklara varför slutbetygen är högre än provbetygen. Låt oss diskutera några av de vanligare förklaringarna utan att för den skull göra anspråk på att det handlar om argument som kan beläggas med empiri från systematiska studier.

a) Läraren har att ta hänsyn till fler mål än de som ingår i nationella provet därför blir slutbetyget högre än provbetyget.

³³ Se appendix för en närmare beskrivning.

Dnr 01-2003:2038

I provmaterialet och i skolans styrdokument betonas att prov inte kan mäta alla mål och att läraren därför har att ta hänsyn till övriga mål vid betygssättningen av eleven. Till de mål som provas i nationella prov ska alltså ytterligare läggas de mål som proven inte prövar.

Om man då till de mål som de nationella proven prövar lägger de ytterligare mål som läraren har att pröva så bör - om man antar att några elever som klarat alla mål på provet inte klarar något eller några av de mål läraren har att pröva - andelen elever som totalt sett inte nått samtliga mål vara *högre* i slutbetyget än på det nationella provet. Man kan alltså knappast förklara de högre slutbetygen med att läraren har att ta hänsyn till *fler* mål än vad som ingår i det nationella provet eftersom det snarare borde höja andelen elever som inte nått samtliga mål. Förklaringen till att slutbetygen är högre än provbetygen måste alltså vara en annan.

b) Läraren har ett bredare underlag än provresultatet för att bedöma elevernas målpuppfyllelse.

En annan förklaring kan vara att läraren har ett mycket bredare underlag och att det visserligen är så att en elev inte råkade klara tillräckligt många mål (poäng) på provet för att bli godkänd, men att läraren i andra sammanhang sett att eleven uppnått målen och att eleven därför i slutändan kan godkännas om han eller hon dessutom uppfyller de mål läraren har att pröva. Det är då inte det förhållandet att läraren ska pröva fler mål som är förklaringen utan förhållandet att läraren har ett mycket *bredare* underlag för att bedöma de mål som proven behandlar. Denna förklaring förefaller rimlig.

Omvänt skulle man kunna tänka sig att läraren ansett att en elev tidigare inte nått de mål som det nationella provet prövar men att han eller hon blir godkänd på det nationella provet och därigenom visar att målen nu är nådda. Det skulle kunna förklara den vanliga kommentaren att nationella proven endast kan höja, inte sänka betyget.

För gymnasieskolan där läraren ofta inte har ett lika brett underlag för sin bedömning är måhända denna förklaring mindre relevant.

c) Slutbetygen är högre därför att särskilda insatser sätts in för elever som inte klarat proven.

I grundskolan ges de olika delproven vid olika tidpunkter under vårterminen i år 9. Om en elev inte klarat ett prov finns då, åtminstone för vissa delprov, tid att åtgärda de kunskapsbrister eleven visat. Genom särskilda åtgärder skulle det då vara möjligt att ge eleven möjlighet att nå målen och få godkänt slutbetyg. Detta förefaller vara en rimlig förklaring när det gäller vissa elever som inte blivit godkända på provet. Däremot torde sådana stödåtgärder knappast sättas in för elever som redan är godkända, eller mer därtill, och denna förklaring är då inte relevant för att förklara varför andelen elever med betyget MVG är större för slutbetyg än provbetyg.

Dnr 01-2003:2038

Huruvida en förklaring om särskilda stödåtgärder har samma relevans för gymnasieskolans elever med IG på provet är svårt att bedöma. För grundskolans del är det väsentligt att få eleverna godkända för att de ska bli behöriga för gymnasieskolans nationella program. Därför ges ofta stödundervisning i form av extraundervisning, sommarskolor etc. Motsvarande aktiviteter förekommer också inom gymnasieskolan men det är svårt att bedöma om omfattningen är lika stor som för grundskolan.

d) Provgruppen och lärarkåren har olika tolkningar av mål och betygskriterier

Att sätta betyg innebär att tolka de nationella kursmålen och betygskriterierna och att ställa dem i relation till olika elevprestationer. Den grupp som bestämmer betygsgränserna på proven består av särskilt utvalda aktiva lärare som i samverkan med provkonstruktörerna tolkar måldokumentet och bedömer provuppgifterna. Det är ingen tvekan om att detta är en kompetent, kunnig och välinformerad grupp och att betygsgränserna fastställs i enlighet med etablerade och vedertagna metoder. Det är dock, som tidigare nämnts, en grupp på cirka 15 personer som gör bedömningen.

Slutbetygen däremot kan sägas vara lärarkårens samlade (om än inte samordnade) bedömning av eleverna och tolkning av måldokumentet. En möjlig förklaring till skillnaden mellan provbetyg och slutbetyg skulle då kunna vara att bedömargruppen genom att ingå i provuppdraget utvecklar en annan (och för matematikens och svenskans del strängare) bedömningspraxis än den samlade lärarkåren.

Det är svårt att bedöma hur trovärdig en sådan förklaring kan vara. Och kanske ännu svårare att avgöra vilken grupp som i så fall gör den ”rätta” tolkningen och bedömningen.

e) Det finns kvar rester av ett grupprelaterat tänkande som gör att andelen som inte når målen inte anses få bli för stor.

En svårbedömd förklaring men det förefaller inte orimligt att tolkningen av målen är beroende av vilken eller vilka elevgrupper man har. De nationella målen och betygskriterierna är skrivna i en generell form som gör att det finns ett betydande tolkningsutrymme. Detta i sin tur gör det naturligt att tolka innebörden i dokumenten utifrån den eller de elevgrupper man är van att arbeta med. Förklaringen förefaller därmed rätt rimlig.

f) Behörighetskraven gör att man inte gärna vill underkänna elever.

Olika studier som Skolverket gjort och information som på annat sätt kommer till Skolverket tyder på att detta är en förklaring som man inte kan bortse ifrån. Det har hänt att lärare har ringt och berättat om sina vändor inför betygssättningen. Man känner sig ensam och utlämnad i en svår process. Även om man egentligen tycker att eleven inte nått alla mål så trycker eleven på om att bli godkänd, och elevens föräldrar gör samma sak. Ofta har inte heller skolledningen något

Dnr 01-2003:2038

intresse av att elever inte blir godkända utan ser helst att andelen elever som inte når målen är så liten som möjligt. Det ser bättre ut i både den kommunala och nationella statistiken. Även studievägledaren uppges ibland uttrycka dubier mot att elever blir underkända eftersom de vet att det individuella programmet i många fall inte fungerar särskilt bra. Det är i sådana fall förståeligt, om än inte försvarbart, att läraren ger vika och sätter det godkända betyg alla vill ha. (Det bör dock betonas att uppgifterna bygger mer på anekdotisk information än på systematiska studier).

För gymnasieskolans del har denna förklaring knappast samma relevans även om det naturligtvis förekommer betygsstress även där. Denna stress är dock mer relaterad till de högre betygen och elever på de studieförberedande programmen. Rädslan för det individuella programmet sätter i grundskolan en särskild press på att få eleven godkänd i de ämnen som ger behörighet till ett nationellt program.

g) Eleverna har arbetat med ett annat stoff än det provet behandlar

Detta torde i huvudsak vara en teoretisk invändning. Enligt ansvarsfördelningen i skolan är det lärare och elever som ska välja stoff och arbetsmetoder. Om valet av stoff är ett annat än det som ingår i nationella provet skulle följden kunna bli att den på provet uppvisade måluppfyllelsen blir sämre än den måluppfyllelse eleverna uppvisat i klassrumsarbetet.

Men frågan är hur vanligt det är att lärare och elever väljer stoff som avviker från det gängse förekommande. Mycket tyder på att läromedlens ställning är stark när det gäller valet av innehåll i undervisningen. Läromedelsförfattare brukar av tradition studera de nationella proven och deras innehåll. Detta får rimligen genomslag i läromedlen och stoffvalet torde därmed i praktiken inte skilja sig särskilt mycket mellan olika skolor. Dessutom är i synnerhet svenska och engelska ämnen med påtaglig progression som inte är särskilt stoffberoende. Matematik är det kanske i högre grad, men å andra sidan är matematik det ämne av de tre där kursplanerna är mest konkreta när det gäller att ange innehåll i kurserna.

Förklaringen är således inte särskilt trovärdig.

h) Proven kommer så tidigt att hela kursen inte hunnit läsas in

I grundskolan är det fyra år mellan målen i år 5 och målen i år 9. Det är då knappast så att tidsbrist är någon avgörande faktor när det gäller skillnaden mellan betyg på det nationella provet och betyg enligt lärarens bedömning. När det däremot gäller gymnasieskolan, där kurserna är betydligt kortare kan tidpunkten för provets genomförande ha större betydelse. Även om flertalet prov ligger i maj och december kan en relativt större del av kurstiden återstå efter provtillfället för gymnasieskolan än för grundskolan. Alternativet kan således ha visst förklaringsvärde när det gäller gymnasieskolan.

h) Frågan är irrelevant, det finns ingen relation

Dnr 01-2003:2038

Denna uppfattning är inte helt ovanlig. Enligt den så är det läraren som har uppdraget att sätta betyg och det är därmed slutbetyget som per definition är det riktiga oberoende av i vilket förhållande det står till provbetyget. Konsekvensen av denna uppfattning blir rimligtvis att det inte finns någon anledning att ha provbetyg. Ett provbetyg skapar då endast förvirring eftersom det uttrycks i samma termer som slutbetyg, men gäller något annat. De borttagna provbetygen i svenska och engelska kan ses som konsekvenser av ett sådant synsätt. Dock finns fortfarande betyg på de tre delprov som ingår i dessa prov och en naturlig fråga blir då hur man ska se på relationen mellan betyget på respektive delprov och slutbetyget i ämnet.

Ingen av de olika relevanta förklaringarna till skillnaderna mellan slutbetyg och provbetyg torde ensam vara tillräcklig. Flera förklaringar verkar mer eller mindre rimliga och det är svårt att med utgångspunkt i gällande styrdokument och föreskrifter entydigt säga att relationen bör vara på ett bestämt sätt. Utan att i det här läget fastlägga vilken den bör vara kan det dock finnas skäl att undra över varför relationen mellan provbetyg och slutbetyg varierar mellan ämnen, mellan skolor, mellan program, men inte mellan skolformer.

Skillnader mellan ämnen

Det framgår klart att ämnena skiljer sig åt. I engelska har ungefär lika många elever högre slutbetyg än provbetyg som tvärtom. Fyra av fem har dock samma betyg. I matematik är avvikelserna störst. Där har cirka 25 procent av grundskolans och 20 procent av gymnasieskolans elever ett högre slutbetyg än provbetyg. Svenskämnet intar en mellanställning.

Varför skiljer sig då ämnena åt? Varför är slutbetygen systematisk högre än provbetygen i matematik och svenska medan de är både lägre och högre i engelska? Har ämnena olika karaktär eller finns det någon annan tänkbar förklaring? Är matematikproven ”svårare” än proven i engelska? Eller kanske annorlunda uttryck är kraven högre i matematik? Kan kraven i olika ämnen överhuvud jämföras?³⁴

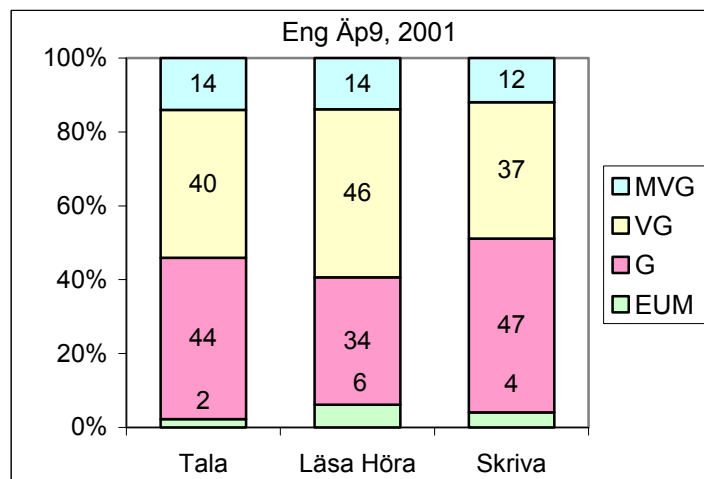
³⁴ Det ska omedelbart sägas att ”lättare” och ”svårare” kan ges lite olika innebörder. En innebörd är empirisk och innebär att elevernas poängfördelning på ett prov avgör om det kan kallas lätt eller svårt. Ett prov där eleverna får höga poäng kallas då lätt och ett prov där eleverna får låga poäng (i relation till vad som är möjligt) kallas svårt. Det säger ingenting om att provet är lätt eller svårt i någon mer absolut mening. I en annan mening kan ”lätt” eller ”svårt” ses i relation till vilken typ av tänkande som krävs, dvs. om det ligger på en lägre eller högre kognitiv nivå. Att lösa problem och analysera komplicerade sammanhang anses då svårare än att t.ex. återge minneskunskaper. Svårighetsgraden bör dock inte blandas samman med arbetsinsatsen. Om man t.ex. säger att det är svårt att läsa franska menar man troligen att det krävs mycket arbete och träning, inte att det skulle vara ett språk som ligger på en annan kognitiv nivå än t.ex. engelska eller svenska.

Dnr 01-2003:2038

Poängjämförelse av matematikprov och ett delprov i engelska

Redovisningen i det här avsnittet bygger på tillgänglig statistik om prov och betyg. Detta innebär naturligtvis en väsentlig reduktion av den information som proven ger. En poängsumma ger inte samma information som en granskning av ett prov och de svar eleven lämnat. Det är väsentligt att ha detta i åtanke så att inte alltför långtgående slutsatser dras. Vissa delprov är inte poängsatta och ger därmed inte underlag för mer detaljerade statistiska jämförelser. Men matematikproven är poängsatta och delprov B i engelska likaså. Kan dessa poängfördelningar bidra med någon förklarande information?

Uppgifterna i delprov B i engelska redovisas med poäng när skolorna redovisar sina provresultat och för matematikens del redovisas poäng för alla delprov. I det senare fallet i form av två olika slag av poäng, så kallade g- och vg-poäng, där g-poäng gäller kunskaper redovisade på "Godkänd"-nivå och vg-poäng kunskaper som anses visa minst "Väl godkänd"-kvaliteter. I den här redovisningen läggs de två poängtyperna samman till en totalpoäng. Matematikpoängen representerar således hela matematikprovet medan poängen i engelska endast gäller det delprov som rör engelsk hör- och läsförståelse (delprov B). Provdelenarna tala (delprov A) och skriva (delprov C) är inte poängsatta utan betygssätts efter en holistisk bedömning baserad på fastlagda kriterier. Figur 43 visar betygsfördelningen för de tre delproven som ingår i provet 2001 i engelska för år 9. Den aktuella poängfördelningen gäller således delprovet "Läsa Höra" i diagrammet.



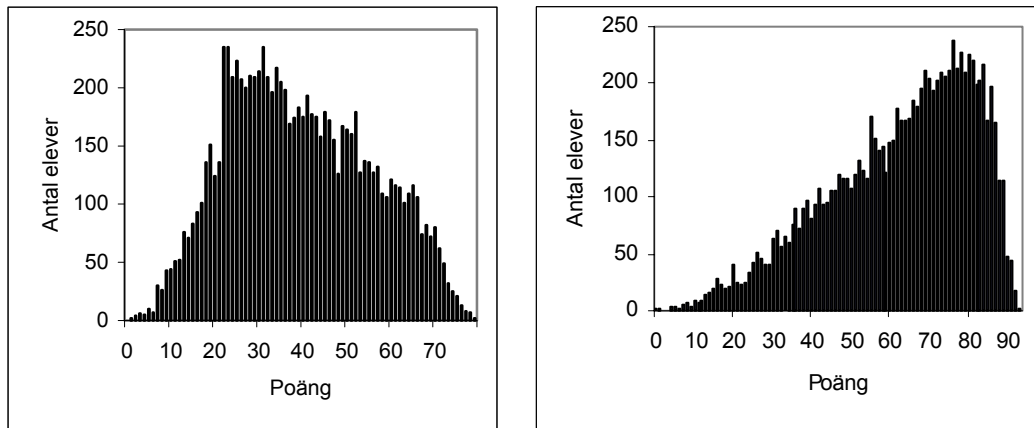
Figur 43. Betygsfördelning för de tre delprov som ingår i ämnesprovet i engelska för år 9 vårterminen 2001³⁵.

³⁵ Det kan synas märkligt att en större andel elever nått målen i skriva (96%) än i läsa/höra (94%). Det beror naturligtvis på hur målen är formulerade. Resultaten ska således inte tolkas i absolut mening, dvs. att man antingen kan eller inte kan läsa, utan hur man läser i relation till målen.

Dnr 01-2003:2038

Man kan konstatera att flest elever (60 procent) har VG eller MVG på det aktuella delprovet, men också att detta prov har största andelen elever som ej nått målen.

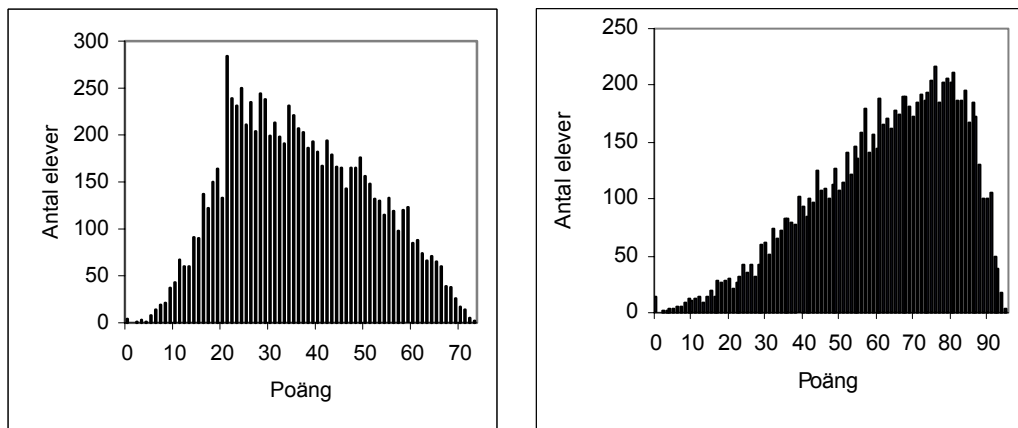
Figur 44 visar poängfördelningen för delprov B i engelska och på matematikprovet. I båda fallen gäller det proven år 2001.



Figur 44. Poängfördelning på provet i matematik för år 9 vt 2001 (till vänster) och för delprov B i ämnesprovet i engelska samma år (till höger i figuren).

Figur 44 visar att poängfördelningarna närmast ser ut som varandras spegelbilder. Fördelningarna är påtagligt sneda åt var sitt håll. Matematikpoängen har sitt typvärde (flest elever) precis vid gränsen för godkänd medan engelskan har sitt nära gränsen för MVG.

Måhända är dessa fördelningar något som råkar gälla för proven 2001 så låt oss också se fördelningarna för 2002 års prov (figur 45).



Figur 45. Poängfördelning på provet i matematik för år 9 vt 2002 (till vänster) och för delprov B i ämnesprovet i engelska samma år (till höger i figuren).

Figur 44 och 45 framstår närmast som kopior av varandra, vilket kan ses som ett uttryck för att provkonstruktörerna lyckas väl med att konstruera likvärdiga (parallella) prov de olika åren.

Vad kan man då säga om att poängfördelningarna är så olika för de två ämnena? Det finns givetvis många tänkbara förklaringar varav en här ska tas upp till diskussion.

Dnr 01-2003:2038

Har ämnena olika karaktär?

Det har tidigare framgått att relationerna mellan provbetyg och slutbetyg skiljer sig påtagligt mellan olika ämnen, främst engelska och matematik. Ovan kunde också konstateras att poängfördelningarna på matematikprovet respektive delprov B i engelska ser påtagligt olika ut. Kan dessa skillnader förklaras med att ämnena är olika i några väsentliga avseenden?

Göran Lindhe, som är docent i pedagogik vid Lärarhögskolan i Stockholm, har på uppdrag av Skolverket gjort en analys av kursplaner och betygskriterier (Linde, 2003). Därvid har han bland annat undersökt matematik och engelska.

Enligt Lindhe finns det påtagliga kunskapsteoretiska skillnader i kursplanerna för respektive ämne. Matematiken beskrivs enligt Lindhe som "ett verktyg för förståelse och med hänvisning till vardagsliv och problemlösning". Däremot anser han för engelskans del att

I engelska är den empiristiska hållningen till hur kunskapsförvärv går till mycket uttalad. Att deduktivt sluta sig till hur en korrekt mening kan vara konstruerad genom användning av en uppsättning regler, är inte det sätt på vilket man lär sig engelska. Det är exponeringen för kommunikativa situationer på engelska som leder till kunskap i engelska. Kunskapen är identisk med kommunikativ förmåga. Språkanalytisk förmåga nämns inte (sid.67).

Om Lindhes analys är korrekt skulle en tolkning kunna vara att matematik i högre grad än engelska utmärkas av krav på analys och förståelse med syfte att lösa olika typer av problem. Engelska å andra sidan skulle domineras av muntlig och skriftlig kommunikation utan några mer uttalade krav på analys eller förståelse av språkliga strukturer och samband. Betygskriterierna i engelska för Väl godkänd och Mycket väl godkänd uttrycks också huvudsakligen i termer av att eleven "använder", "tillgodogör sig", "uttrycker sig" etc., och inte lika påtagligt som i andra ämnen i andra ämnen, att eleven "reflekterar", "diskuterar", "analyserar" osv.

Om man relaterar det sagda till någon kognitiv taxonomi (t.ex. Blooms) så brukar reflektion, analys, problemlösning och liknande förmågor anses ligga på en högre kognitiv nivå än "tillägnandet" av olika färdigheter av det slag som kursplanerna i engelska kan anses ge uttryck för.

Om de förda resonemangen är riktiga och om samma tankegångar som uttryck i kursplaner och betygskriterier genomsyrar de nationella proven, vilket är rimligt eftersom proven bygger på noggrann läsning av dessa dokument, förefaller det inte orimligt att matematikproven med sina krav på högre kognitiva förmågor framstår som "svårare" än proven i engelska. Därmed skulle de skillnader som föreligger i poängfördelningarna för matematikprovet och provet i engelska kunna förklaras med kursplanernas målformuleringar och betygskriterier.

Dnr 01-2003:2038

I det mer krävande matematikämnet samlas resultaten i den nedre delen av skalan medan resultaten i engelska till följd av lägre krav på reflektion, analys mm., grupperar sig kring betygskalans övre gränser.

Om Lindhes analys är korrekt och korrekt tolkad kan alltså de olika poängfördelningarna för engelska och matematik ses som en följd av att kursplanernas mål och betygskriterier i respektive ämne har skilda kunskapsteoretiska utgångspunkter. Skillnaderna i poängfördelningar ska således inte ses som uttryck för att proven inte är välkonstruerade, utan snarare att de väl speglar styrdokumentet. I ett målrelaterat system är inte ambitionen att konstruera prov som differentierar provdeltagarna över en poängskala utan att pröva om de nått målen för kursen eller ämnet. Och om alla gjort det i enlighet med kriterierna för t.ex. MVG så ska alla ha MVG, även om fördelningen därmed blir sned.³⁶

Om man av någon anledning anser att de aktuella poängfördelningarna är olämpliga skulle den logiska följden av ovanstående resonemang bli att kursplaner och betygskriterier i engelska bör revideras mot att mer betona reflektion, analys, argumentation och problemlösning, medan följden för matematikens del antingen innebär ökad satsning på matematikundervisningen, eller mildring av kraven i mål och betygskriterier. Detta är frågor av utbildningspolitisk natur.

Men även om ämnenas skilda karaktärer enligt kursplanerna förklarar olika poängfördelningar så förklarar de inte varför relationen mellan provbetygen och lärarnas slutbetyg skiljer sig mellan ämnena. Samma kursplaner och betygskriterier ligger ju till grund både för lärarnas betygssättning och för provkonstruktörernas bedömningar. Om matematik på grund av kursplanernas målformuleringar leder till svåra prov så borde rimligen lärarnas bedömningar utifrån samma kursplaner ligga på samma nivå. Som de tidigare resultaten visat varierar dock förhållandet mellan provbetyg och slutbetyg en hel del mellan skolor. För engelskans del ligger variationen nära ”nollinjen” medan den för matematikens ligger på en systematiskt högre nivå. Även om det som tidigare framhållits kan finnas rimliga förklaringar till variationen mellan skolor är det svårt att förklara varför matematiken varierar runt en generellt större höjning av slutbetyget.

Möjligen, och det förefaller rimligt, är det så att man som lärare är mer benägen att höja slutbetyget när provresultaten samlas vid (eller under) gränsen för godkänd än när man uppfattar att betygsgränserna på provet är ganska lågt satta, vilket fler lärare i engelska än i matematik gör enligt enkätsvar från lärare som deltagit i proven.

³⁶ Linde anser efter sin analys att kriterierna för olika betyg i såväl matematik som engelska (kurs A) ger uttryck för skilda kunskapskvaliteter i enlighet med betygssystemets intentioner.

Dnr 01-2003:2038

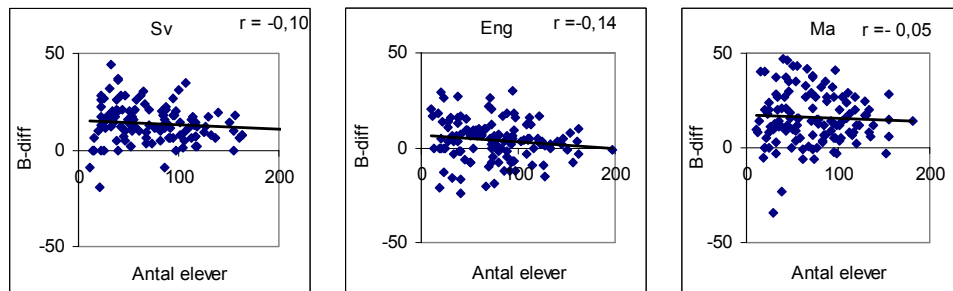
För svenskämnet finns inga poängfördelningar tillgängliga och det ingår inte heller i Lindhes analys. Det har därför inte kommenterats i det här sammanhanget.

Skillnader mellan olika skolor

Som framgått kan skillnaderna mellan olika skolor vara påtagliga när det gäller relationen provbetyg – slutbetyg. Frågan är om man kan hitta något i det statistiska underlaget som förklarar detta.

Skolstorlekens betydelse

Ett tänkbart samband kunde vara att det är skillnad mellan små och stora skolor. Figur 46 visar sambandet mellan ”B-diff” (andelen elever som har ett slutbetyg som ligger ett steg högre än provbetyget) och skolstorlek uttryckt i antalet elever.



Figur 46. Samband mellan betygskillnad (slutbetyg – provbetyg) och skolstorlek.

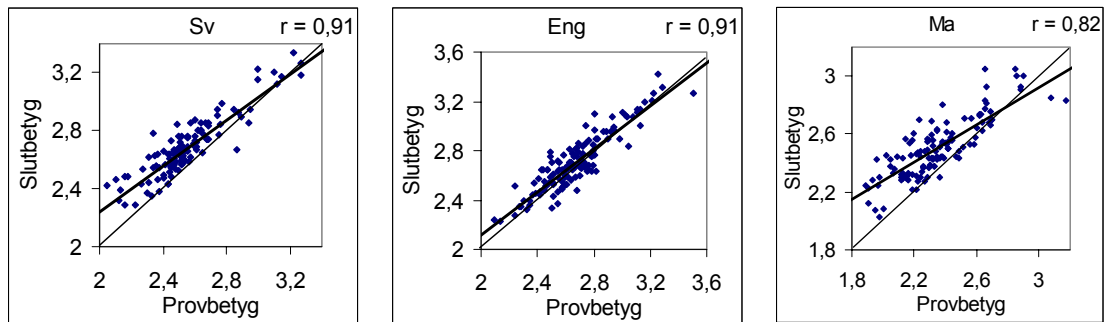
Av figur 46 framgår att antalet elever vid skolan inte systematiskt samvarierar med skillnaden mellan slutbetyg och provbetyg. Vad man kan se är att variationen är större för små skolor än för stora (fler små skolor med stora skillnader i båda riktningarna). Detta är rimligt med tanke på att slumpinflytandet bör vara större för små skolor, liksom omvänt att de stora skolorna grupperar sig närmare trendlinjen.

Samband mellan skolmedelvärden för provbetyg och slutbetyg

En betraktelse av sambanden mellan skolornas medelvärden för provbetyg och slutbetyg kan leda till en del intressanta iakttagelser (figur 47).³⁷

³⁷ Den rättrådige kan här givetvis invända att det inte är korrekt att tala om medelvärden när det gäller betyg som är baserade på kvalitativa bedömningar av måloppfyllelse. Låt oss dock temporärt och approximativt anta att det kan vara meningsfullt att tala om medelvärden för betyg. I det här fallet har betygen tilldelats värden enligt IG eller EUM=1, G=2, VG=3 och MVG=4. Den asymmetriska skala som fastlagts i olika föreskrifter (IG/EUM=0, G=10, VG=15 och MVG=20) används således inte i det här sammanhanget.

Dnr 01-2003:2038



Figur 47. Samband mellan skolmedelvärde för provbetyg och slutbetyg år 9, vt 2001.

Man kan notera att samstämmigheten är störst för engelska och svenska (korrelationen = 0,91). Vidare kan man notera att trendlinjen i relation till diagonalen har en mindre lutning för matematik och svenska än för engelska. Det innebär att för dessa ämnen är skillnaden mellan slutbetyg och provbetyg större för de skolor som har låga provresultat än för de som har höga, vilket förefaller rimligt.

Man kan slutligen notera att för engelskan ligger skolorna samlade i närheten av diagonalen, med ungefär lika många skolor över som under vilket innebär att provbetyg och slutbetyg ligger nära varandra (vilket också konstaterats tidigare). För svenska och matematik ligger nästan alla punkter ovanför diagonalen vilket innebär att dessa skolor har högre slutbetyg än provbetyg, dvs. samma resultat som erhöles tidigare.

Den slutsats man möjligen kan dra är att det tycks föreligga en tendens som kan liknas vid en "regression mot medelvärdet". Skolor med låga provresultat tenderar således att höja slutbetygen mest medan skolor som presterat bra provresultat har en viss tendens att ge lägre slutbetyg än provbetyg, eller i varje fall att inte höja dem lika mycket.

Samtidigt förefaller det inte orimligt att det främst är skolor som lyckats dåligt på provet som har den största ökningen i slutbetygen. Dessa skolor torde ha ett större intresse av att sätta in stödåtgärder än skolor som lyckats bra på provet.

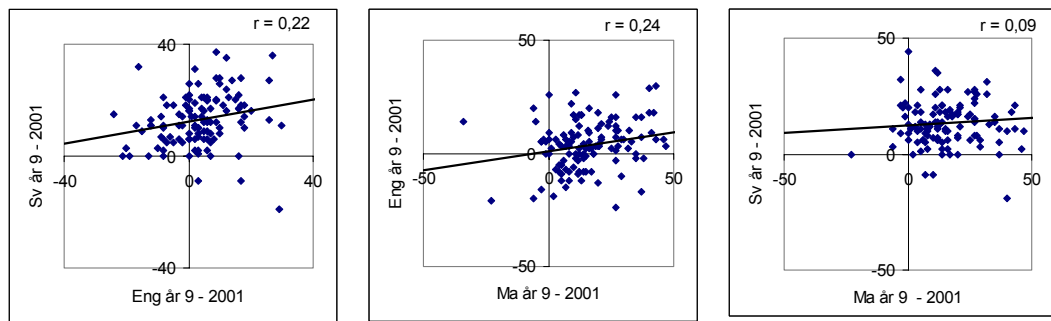
Finns "högbetygsskolor" i urvalet?

Ibland talas det om att vissa skolor är "högbetygsskolor" och andra "lågbetygsskolor". Om så vore fallet skulle det kunna förklara varför skillnaden mellan provbetyg och slutbetyg varierar mellan olika skolor. Om man betraktar provbetyget som norm skulle då högbetygsskolor genomgående ha stora skillnader mellan slutbetyg och provbetyg medan lågbetygsskolor skulle ha små eller t.o.m. negativa skillnader. Om beteckningarna högbetygsskola och lågbetygsskola ska ha fog för sig bör det då vara så att samma skolor utmärker sig genom stora respektive små skillnader i alla provämnen, dvs. korrelationen

Dnr 01-2003:2038

mellan andel elever med höjda betyg i olika ämnen bör vara positiv. Låt oss därför undersöka de skolor som ingår i urvalet.

Figur 48 visar sambandet mellan respektive ämnen när det gäller skillnader mellan slutbetyg och provbetyg för de skolor som ingår i urvalet.



Figur 48. Samband mellan andel elever med ett steg högre slutbetyg än provbetyg för de olika provämnena för de skolor som ingår i urvalet vårterminen 2001.

Av figuren kan man konstatera att det finns ett svagt positivt samband för engelska och svenska (korrelation = 0,22), dvs. det tycks finnas en viss tendens att vid somliga skolor är skillnaderna mellan slutbetyg och provbetyg större i både svenska och engelska än i vissa andra skolor.

Även för engelska och matematik finns ett visst positivt samband (korrelation = 0,24), vilket alltså indikerar att det finns en tendens att vissa skolor har större skillnader än andra mellan slutbetyg och provbetyg i båda ämnena.

När det gäller matematik och svenska är dock sambandet i stort sett försumbart (korrelation = 0,09) och för dessa ämnen finns det knappast fog för att tala om att vissa skolor har särskilt stora respektive små skillnader mellan slutbetyg och provbetyg i båda ämnena samtidigt. Som figur en visar kan skillnaderna vara stora och positiva i det ena ämnet och obefintliga eller negativa i det andra.

Man kan möjligen sträcka sig till att säga att det finns visst stöd för att tala om att vissa skolor tenderar att ha större skillnader mellan slutbetyg och provbetyg än andra. Det är dock förhållandevis svaga samband och några mer långtgående slutsatser kan därför inte dras. Det förefaller således snarare som om man (under antagandet att provbetyget är norm) kan tala om högbetygsskolor, eller lågbetygsskolor, när det gäller enskilda ämnen men knappast generellt för alla tre ämnen.

Den fråga som infinner sig men som inte kan besvaras på basis av den här gjorda redovisningen är: Vad innebär det för elever att gå i en skola med stor andel höjda betyg i relation till att gå i en skola med få höjda betyg?

Dnr 01-2003:2038

Skillnader mellan olika program

Som tidigare nämnts är bilden mer komplicerad för gymnasieskolans del än för grundskolan. Bortfallet är därtill större vilket gör underlaget osäkrare. Det som ändå är slående är att skillnaden i andel elever med slutbetyg som ligger ett steg högre än provbetyg är förhållandevis liten och osystematisk mellan olika program när det gäller svenska och engelska. Däremot är det för matematikens del påtagligt att det framför allt är elever vid yrkesinriktade programmen som erhåller högre slutbetyg än provbetyg. Möjligen kunde då skillnader mellan olika gymnasieskolor, främst när det gäller matematik, kunna förklaras med hur stor andel elever som finns på yrkesinriktade program.

Skillnader mellan grundskola och gymnasieskola

Det som är slående vid en jämförelse mellan grundskola och gymnasieskola är hur likartad bilden är. För båda skolformerna gäller att överensstämmelsen mellan provbetyg och slutbetyg är störst i engelska och minst i matematik. Möjligen kan man förklara detta med att, i varje fall när det gäller engelska och matematik, att grundskolans kursplan och A-kursen på gymnasiet till stor del omfattar samma innehåll.

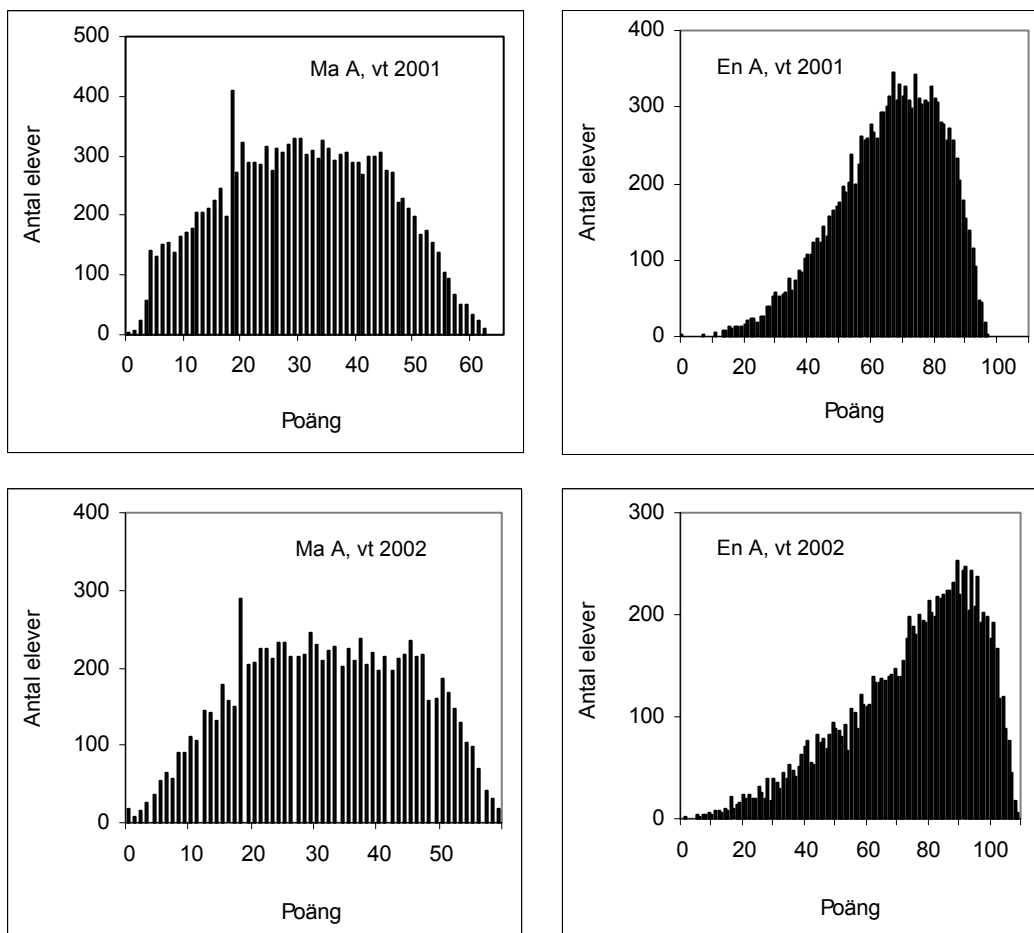
En annan möjlig faktor kunde vara att både grundskolans prov och de aktuella kärnämnesproven konstrueras vid samma institutioner (även om det inte alltid är samma personer som arbetar med respektive prov), och att detta leder till ett visst sätt att tolka och förstå ämnet. Detta behöver givetvis inte vara något negativt, tvärtom ligger det ett värde i att det finns en kontinuitet i progressionen vid övergången mellan de olika skolformerna.

Men en följd av den enhetliga miljön skulle kunna vara att den standard som sätts på grundskolan också överförs till gymnasieskolans kärnämnesprov. På så sätt kommer de förhållandevis "lätta" engelskproven i grundskolan att leva vidare i kärnämneskursen engelska A och de förhållandevis "svåra" matematikproven i år 9 att lägga grunden för provet för gymnasiekursen matematik A. Ett gymnasieprov kan ju av legitimitetsskäl knappast tillåtas bli mindre krävande än grundskolans prov i samma ämne. Ett "svårt" grundskoleprov genererar således ett "svårt" gymnasieprov till i princip samma elever som ett år tidigare gjorde grundskoleprovet (med undantag för de som ej blivit behöriga för ett nationellt program). Det förefaller då inte så förvånande att svarsmönstret är likartat i matematik.

I den tidigare nämnda studien av Lindhe över kursplaner och betygskriterier konstateras att "Skiljelinjen går främst mellan ämnen inte mellan stadier och åldersgrupper i skolsystemet". Utifrån Lindhes analys är det med andra ord inte överraskande att mönstret för grundskolan och gymnasieskolan är detsamma. Skillnaderna mellan ämnena skulle då främst vara en följd av de olikheter i kunskapssyn som präglar de olika ämnenas kursplaner.

Dnr 01-2003:2038

En granskning av poängfördelningarna för kärnämneskurserna matematik A och engelska A (delprov B läs- och hörförståelse) förefaller vid en jämförelse med motsvarande poängfördelningar för grundskolan ge stöd för den tidigare framförda tolkningen av Lindes analys.



Figur 49. Poängfördelningar för matematik A vt 2001 (överst till vänster) och för läs- och hörförståelsedelen i provet för engelska A samma termin (överst till höger). De undre diagrammen visar motsvarande poängfördelningar vt 2002.

Man kan notera att poängfördelningarna i engelska överensstämmer mycket väl med motsvarande poängfördelningar på proven i år 9.

För matematikens del är dock fördelningen mindre sned för kurs A än för år 9. Det som är anmärkningsvärt i figur 49 för matematikens del är hur väsentligt det tycks vara för många lärare att eleverna klarar godkänt på provet. För båda matematikproven gäller att det finns en påtaglig anhopning av elever vid gränsen för godkänt. En sådan fördelning är naturligtvis inte sannolik utan tyder på att poängsättningen påverkats av var godkändgränsen legat. Detta tyder i sin tur på att provresultaten tilldelas stor betydelse. Som tidigare framhållits betonar skolans styrdokument att det är läraren som ska sätta betyget på basis av sitt samlade underlag om elevens prestationer och att

Dnr 01-2003:2038

de nationella proven endast utgör en del av detta samlade underlag. De nationella proven ska inte betraktas som examensprov. Poängfördelningen ovan indikerar att den informationen inte till fullo nått fram.

Relationen mellan provbetyg och slutbetyg generellt

Vad kan man då dra för slutsatser av den genomförda redovisningen?

De nationella proven har som tidigare beskrivits olika syften, varav ett – och kanske det viktigaste – är att stödja lärarna i en likvärdig bedömning och rättvis betygssättning. Gör proven det och i så fall på vilket? Lärarnas och rektorernas uppfattning i dessa frågor har redovisats i kapitel 2. Här kommer några mer principiella frågor att diskuteras.

Man kan tänka sig att de nationella proven kan fungera betygsstödjande på minst tre olika sätt,

- 1) som mått på den enskilda elevens måluppfyllelse,
- 2) som mått på den egna gruppens grad av måluppfyllelse,
- 3) som förebildligt stödmaterial för lärare och lärarlag.

Kategorierna är naturligtvis inbördes överlappande men kan ändå tjäna som underlag för en diskussion.

Prov som mått på den enskilda elevens måluppfyllelse

Provresultat som mått på en individs kunnande eller förmåga har en rätt lång tradition. De gamla real- och studentskrivningarna var ofta av avgörande betydelse för om eleverna skulle godkännas eller inte. Proven fungerade som examensprov. Denna form av examinerande prov har inget stöd i dagens svenska läroplaner. De försvann ur det svenska skolsystemet när grundskolan och den reformerade gymnasieskolan infördes i mitten av 1960-talet. Dessa prov var varken grupprelaterade eller målrelaterade utan snarare absoluta. Några tydliga angivelser om vad elever skulle kunna för att få vissa betyg fanns inte utan om man t.ex. hade tre av åtta uppgifter rätt var man godkänd. Provresultatet gav en certifiering (till fortsatta studier) men sa i övrigt ingenting om vad man kunde eller hur man förhöll sig till andra elever.

Idag finns en annan medvetenhet om vad prov kan visa och vad de inte kan visa. När det gäller att pröva enskilda individers kunnande måste provresultat tolkas med försiktighet. I *Standards for educational and psychological testing*³⁸, det mest genomarbetade kvalitets-säkringssystem som existerar för provkonstruktion och provanvändning skrivs (Standard 13.7, sid 146):

³⁸ American Educational Research Association m.fl. (1999)

Dnr 01-2003:2038

In educational settings, a decision or characterization that will have major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision.

Att basera viktiga beslut - t.ex. att sätta betyg - på enbart ett prov står inte bara i strid mot läroplanens föreskrifter, det står också i strid mot professionens egna kvalitetskriterier.

Frågan är dock om föreställningen om examensprov idag har försvunnit ur "det allmänna medvetandet" om vad ett nationellt prov är. Skolverkets kvalitetsgranskning om betyg år 2000 och frågor som både föräldrar och elever riktar till Skolverket tyder på att det bland många vuxna, och även bland lärare, fortfarande finns mer eller mindre starka föreställningar om att de nationella proven har en sorts examinerande funktion. Synpunkter av typen "Man kan inte bli godkänd om man inte har godkänt på det nationella provet." och "Är man godkänd på det nationella provet är man godkänd på kursen." förekommer. Den kanske vanligaste synpunkten är dock att "Det nationella provet kan bara höja inte sänka betyget."

Oberoende av vilken kommentar man väljer så ger den uttryck för att det nationella provet har en särskild betydelse för den enskilda eleven, dvs. att det i någon mening har karaktär av examensprov. Även de tidigare redovisade poängfördelningarna i matematik indikerar att provresultaten tilldelas en speciell betydelse.

Naturligtvis bör resultatet från de nationella proven vägas in i bedömningen när en elev ska betygsättas, men det är viktigt att hela bilden av elevens prestationer beaktas och att provresultatet inte tilldelas alltför stor betydelse. Hur stor betydelse det ska ges är svårt att ange på något mer precist sätt. Det är och måste vara en bedömning som görs av den som har tillgång till hela bilden av elevens prestationer.

Prov som mått på den enskilda gruppens måluppfyllelse

De standardprov och centrala prov som efterträdde examensproven på 1960-talet hade ett uttalat syfte att utgöra mått på den egna gruppens prestationer. Betygssystemet var grupprelaterat och provens syfte var att på ett så bra sätt som möjligt för läraren ge upplysning om hur hans eller hennes klass förhöll sig i förhållande till övriga elever i landet som läste samma kurs. Proven skulle ge nivån för den egna klassen men även en bild av betygsfördelningen. Däremot framgick det av anvisningarna att betygen för de enskilda eleverna inte behövde sammanfalla med betygen på proven (även om det givetvis var det vanligaste även då).

Trots de tydliga anvisningarna om att proven var en form av gruppbedömningsprov uppfattades de centrala proven av många, kanske flertalet, som särskilt viktiga även för den enskilda elevens betyg och karaktären av examensprov var rätt påtaglig.

Dnr 01-2003:2038

Dagens nationella prov är inte lika tydliga. De ska vara ett stöd för läraren i hans eller hennes betygssättning, men på vilket sätt? Om betygsstödet främst skulle gälla att bedöma nivån på den egna gruppen, men inte på den enskilda eleven, skulle ett - måhända något radikalt - alternativ vara att låta eleverna genomföra provet anonymt. Resultatet skulle ge en bild av gruppens måluppfyllelse och betygsfördelning och det skulle sedan vara lärarens uppgift att betygssätta de enskilda eleverna så att gruppens måluppfyllelse (betyg) överensstämmer inom rimliga gränser med det som provresultatet utvisar. Eleverna skulle fortfarande vara motiverade av att göra sitt bästa eftersom gruppens resultat återverkar på bedömningen av den enskilda eleven samtidigt som risken för att provet uppfattas som examensprov minskar. Resultaten skulle därtill vara tillgängliga för nationell publicering på samma sätt som nu. Och antagligen skulle det bli mindre angeläget att försöka komma över prov i förtid, vilket skulle minska provsystemets sårbarhet.

Man kan dock misstänka att tanken på anonymt genomförda nationella prov skulle väcka starkt motstånd, inte minst från dem som uppfattar nationella prov som, om inte ett examensprov, så i varje fall ett nödvändigt underlag för att bedöma den individuella eleven. Om man emellertid skulle vilja betona att provens främsta uppgift är att stödja läraren vid bedömningen av den egna gruppens allmänna nivå, och samtidigt motverka att de uppfattas som examensprov, är det ett konsekvent tillvägagångssätt. På grundskolan torde ett sådant tillvägagångssätt vara mer oproblematiskt än på gymnasieskolan där kurserna ofta är kortare och lärarnas bedömningsunderlag mindre omfattande.

Prov som förebildligt stödmaterial

Som nämnts i andra sammanhang ger det nuvarande styrsystemet lärare och elever ansvaret för att själva välja vilket stoff de vill arbeta med och hur det arbetet ska genomföras. Detta är ett uttryck för den så kallade deltagande målstyrningen. I ett sådant system innebär ett nationellt prov med anspråk på att ge en korrekt bild av en enskild grups kunnande en anomali. Ett prov måste rimligen handla om något och om då en viss grupp valt att arbeta med ett helt annat stoff än det som ingår i provet är de eleverna rimligtvis missgynnade i relation till en grupp som råkat välja samma eller ett närliggande stoff.

Stoffvalets betydelse kan givetvis variera mellan olika ämnen och de provinstitutioner som konstruerar proven tillhandahåller delvis material som ska kunna utgöra gemensamt underlag för provdeltagarna. Trots detta kan man inte med säkerhet utgå från att alla grupper ges likvärdiga förutsättningar att visa sin måluppfyllelse på de nationella proven.

Dnr 01-2003:2038

En konsekvens av detta blir att provet kanske inte har sin viktigaste funktion genom de resultat eleverna visar på proven utan genom den information det ger läraren om hur måluppfyllelse av olika kvalitet kan prövas och bedömas. Stödet skulle då ligga i att proven, och framför allt de utförliga kommentarerna och bedömningsanvisningarna till uppgifterna, utgör förebilder för lärarens *övriga* bedömningsarbete. Man skulle kanske kunna tala om en sorts likvärdighet genom förebildlighet. En text är dock inte i sig tillräckligt utan det är förstås också viktigt att lärare diskuterar uppgifter och bedömningsanvisningar med sina kollegor. Och helst borde denna diskussion föras på nationell nivå.

Prov i ett provsystem som avses fungera betygsstödjande på detta sätt behöver egentligen inte ens genomföras av eleverna. Proven riktar sig inte primärt till eleverna utan till lärarna. Det är en form av kommentarmaterial till kursplaner och betygskriterier där olika exempel på uppgifter och elevlösningar diskuteras och bedöms utifrån målbeskrivningar och betygskriterier. Med detta som grund förväntas lärarna utveckla en form av kollektiv konsensus om hur elevers arbete på ett likvärdigt sätt ska bedömas i relation till mål och betygskriterier.

Den här synen på nationella prov har stöd i skolans styrsystem så som det formulerats i ansvarsproposition och läroplaner. Där betonas att det är lärarna som väljer stoff och arbetsmetoder och lärarna som ska sätta betygen varvid de ska bedöma elevernas samtliga prestationer. För att likvärdighet ska uppnås med denna tolkning av de nationella provens roll borde uppgifter och bedömningar av elevsvar diskuteras offentligt inom lärarkollektivet. I det avseendet utgör de nuvarande sekretessbestämmelserna ett väsentligt hinder.

En konsekvent tillämpning av denna syn på provens roll i skolsystemet skulle vara att proven var frivilliga. De skulle inte heller vara sekretessbelagda eftersom de i första hand kan ses som kommentarmaterial och exempelsamlingar riktade till lärare (och elever, föräldrar m.fl. intresserade). Provbetyg kunde givetvis tillhandahållas som underlag för diskussion och som exempel på hur elevresultat kan relateras till mål och betygskriterier.

Med ett provsystem som ovanstående skulle nya prov inte behöva konstrueras varje år, vilket skulle ge utrymme för framställning av prov, dvs. kommentarmaterial och exempelsamlingar om bedömning och betygssättning, i flera ämnen än de nuvarande. Man kan dock misstänka att ett sådant provsystem skulle uppfattas som kontroversiellt förutom att det skulle få betydande konsekvenser för de institutioner som konstruerar nuvarande prov.

Dnr 01-2003:2038

Utredningsgruppens bedömning

Ett syfte med de nationella proven är att bidra till en likvärdig och rättvis betygssättning. Relationen provbetyg och slutbetyg är dock komplicerad och svårtolkad. Det är inte möjligt utan kännedom om de lokala förhållandena att uttala sig om vilken relationen bör vara. Enligt nuvarande ansvarsfördelning för skolan är det lärare och elever som väljer arbetsmetoder och innehåll i kurserna. I vilken utsträckning dessa val överensstämmer med de val som provgrupperna gjort för de nationella provens del kan inte avgöras utifrån det statistiska underlaget. Slutsatsen blir att inga slutsatser kan dras enbart på basis av det statistiska underlaget. Det är således inte heller möjligt att av de statistiska jämförelserna dra några slutsatser om i vilken utsträckning de nationella proven fyller sitt syfte att stödja en likvärdig och rättvis betygssättning.

Dnr 01-2003:2038

ETT NYTT PROVSYSTEM?

4. Läroplaner och provsystem

Ett kunskapsutvärderingssystem eller provsystem kan inte förstås isolerat utan måste bedömas i relation till det styrsystem inom vilket det verkar. Vid en diskussion av det nuvarande provsystemet är det därför angeläget att ge en lite utförligare beskrivning av de svenska styrdokumenterna och deras utveckling fram till idag samt att försöka identifiera kunskapsbedömningens roll under olika läroplaner.³⁹

Som underlag för analysen används en nyutkommen norsk text som innehåller en intressant typologi.⁴⁰ Den kan på ett konstruktivt sätt tillämpas på det sena nittonhundratalets svenska läroplaner. Den norska studien intresserar sig dock främst för de nu gällande nordiska läroplanerna (till vilka de även räknar kursplaner) medan tillämpningen på de tidigare svenska läroplanerna är en utvidgning inspirerad av den norska studien.

Olika typer av läroplaner

I den norska studien görs en jämförande analys av de nu gällande läroplanerna i Danmark, Norge, Finland och Sverige (Sivesind m.fl., 2003).

Sivesind m.fl. beskriver fyra olika typer av läroplaner som de tyckt sig kunna identifiera. Kategorierna är givetvis inte varandra uteslutande utan det finns gråzoner där likartade drag går igen. De är dock intressanta som verktyg för att analysera t.ex. det svenska skolsystemet under några olika läroplaner. Kategorierna är också intressanta för att de är kopplade till olika typer av instrument för betygssättning, kunskapsbedömning och utvärdering.

1. Läroplanen som ett formellt reglerande dokument

För läroplanstyp 1 anges tyska läroplaner som exempel. Läroplanen är en föreskrivande lagtext som i huvudsak fastställs av jurister. Den preciserar de kunskaper som ska förmedlas i skolan, men anger inte hur undervisningen ska organiseras och genomföras. Tillsynen fokuseras på att undersöka om undervisningens uppläggning är sådan att det täcker det innehåll som föreskrivs i läroplanen. (Däremot går den inte ut på att undersöka om eleverna tillägnat sig detta innehåll). Inget nordiskt land har enligt Sivesind m.fl. en sådan läroplan i dag.

³⁹ För en utförlig redovisning av de läroplaner som föregick Lpo 94 och Lpf 94 hänvisas till Marklund (1987).

⁴⁰ Det finns naturligtvis andra sätt att kategorisera läroplaner. T.ex. talar Lundgren (1979) om olika läroplanskoder, Englund (1986) om konceptioner och Vallberg Roth (2001) om könsdidaktiska koder.

Dnr 01-2003:2038

2. Läroplanen som ett auktoriserande och innehållsbeskrivande dokument

Läroplanstyp 2 har likheter med den föregående i att den på central nivå anger premisserna för vilket innehåll som ska finnas i utbildningen. De skrivs inte längre av enbart jurister utan även pedagoger deltar. Skillnaden mot typ 1 är att angivelserna om innehållet i undervisningen främst gäller mer grundläggande principer om ämnets struktur och metodiken för undervisning i ämnet, liksom angivelser om tvärvetenskapliga kopplingar mellan ämnen. En läroplan av detta slag kan också innehålla principer för urval av innehåll utifrån ämnesmässiga, didaktiska och praktisk-pedagogiska överväganden.

Att denna läroplan anger vida gränser för ämnena får till följd att den inte innebär samma tydliga regelstyrning som föregående typ. Den blir därmed inte heller lika tydlig som underlag för tillsyn (förvaltningsmässig/juridisk kontroll). Läroplaner av detta slag förutsätter professionella lärare som kan fylla de tämligen vida ämnesramarna med ett innehåll. Med hjälp av denna typ av läroplaner kan utbildningsreformer implementeras. Dock anger inte heller denna typ av läroplaner vad eleverna ska kunna. Fokus ligger på lärarnas undervisning, inte på undervisningens resultat.

3. Läroplanen som ett politiskt normerande instrument

Läroplanstyp 3 skiljer sig från de föregående genom att de inte enbart är föreskrifter utan även kan ses som policydokument. De innehåller till betydande del normerande utsagor av politiskt och värderande slag. Här ingår också målbeskrivningar och målstyrningstanken är en viktig utgångspunkt. Forskare och forskargrupper har en viktig roll i läroplansarbetet och aktuell pedagogisk och psykologisk forskning får genomslag. Målformuleringarna är generellt hållna för att säkra den pedagogiska praktikens autonomi.

Uppföljningen i system som utgår från denna typ av läroplan är en annan än för de två föregående typerna där tillsyn av lagföljden är det primära. Enligt Sivesind m.fl. karaktäriseras typ 3 av att lokalt och regionalt utvecklingsarbete är en viktig förutsättning för ett lyckat resultat. Decentraliseringen kräver att alla berörda parter deltar i och tar ansvar för att målen uppnås och för att måluppfyllelsen prövas. Kvalitetsgranskningen görs lokalt och regionalt. Staten har därmed delegerat ansvaret för och kontrollen av måluppfyllelsen till skolans lokala huvudmän. Målstyrningen innebär dessutom att fokus i läroplanen förflyttats från lärarnas undervisning till elevernas resultat.

Målstyrningstanken började, enligt författarna av rapporten, framträda på 1980-talet och de anger att den empirisk-pedagogiska forskning som bedrevs i Sverige lade grunden för den typ av uppföljning som utmärker läroplaner av typ 3.

Dnr 01-2003:2038

4. Läroplanen som ett standardiserande dokument för central utvärdering (evaluering) av utbildningen

En decentraliseringstanke, som bygger på lokalt utvecklingsarbete och kvalitetskontroll, reser frågor om vad som egentligen styr skolverksamheten och om de centrala myndigheterna får tillräckligt tydlig information om det nationella skolsystemet och dess resultat. Ett sätt att söka svar är att införa centrala utvärderingar som på ett systematiskt sätt ger information om utfallet av skolans verksamhet i relation till nationellt fastställda kunskapsmål. Med detta behov för ögonen har begreppet ”standards” lanserats.

En sådan typ av läroplan (läroplanstyp 4) är enligt Sivesind m.fl. ny för Nordeuropa, inklusive Norden. Externa produktkontroller i form av censorer vid studentexamen, inträdesprövningar, konfirmationsförhör etc. har en lång tradition, men sådana har aldrig legat till grund för central utvärdering av systemet. Däremot har denna typ av evalueringar en lång historia i främst USA. ”The standards movement”, som fått stor betydelse för utvecklingen i USA och övriga anglosaxiska länder startade redan på 1980-talet, och är starkt knutet till olika system för kunskapsutvärdering.⁴¹

En passus i Sivesinds m.fl. rapport som har påtaglig relevans i det här sammanhanget är följande

Spørsmålet er om læreplanen skal ha en overordnet legitim status for hvilke evalueringer som gjennomføres, hvordan de gjennomføres og hva som offentliggjøres. Dersom læreplanen ikke stiller krav til disse punktene, vil evalueringen selv få en legitimerende og standardiserende funksjon. Dermed blir evalueringen et styringsvirkemiddel som i seg selv kan stå i et konkurrerende forhold til læreplanen. (sid. 7)

Sivesind m.fl. framhåller att denna typ av styrsystem blir komplicerade och att de kräver noggrant övervägande mellan olika intressen. Dilemmat är att skapa ett system som samtidigt styr både undervisning och evaluering. Risken är att evalueringen blir det som i realiteten styr undervisningen.

Denna fjärde och sista typ av läroplaner, enligt Sivesinds m.fl. kategorisering, har inte status som förordning och innehåller varken föreskrifter eller vägledning om ett givet undervisningsinnehåll. De standards som anger vad eleverna ska kunna kommer dock naturligtvis att uppfattas som, om inte föreskrivande, så i varje fall vägledande. Även om denna läroplanstyp kan uttrycka politiska ambitioner så är den först och främst skriven för att kunna tjäna som underlag för central utvärdering av skolans arbete, främst med avseende på kunskapsmålen .

⁴¹ Se appendix.

Dnr 01-2003:2038

Vilken läroplanstyp har Sverige idag?

Sivesind m.fl. placerar Sverige i läroplanstyp 3, den politisk-normerande. Den starka betoningen av värdegrunden ger stöd för en sådan kategorisering. Kategorierna är dock, som nämnts, inte varandra uteslutande och varje land har större eller mindre inslag av flera läroplanstyper. Således kan man säga att Sverige har inslag av alla typer även om typ 3 dominerar. Typ 1 kan sägas förekomma i samband med tillsynsärenden där lagföljden hos skolhuvudmannen prövas. Kursplanerna kan i sin tur anses ha tydliga inslag av typ 2 genom sina avsnitt om ämnenas karaktär och syfte och de krav som ställs på professionell kompetens för att konkretisera kursplanernas generella ämnesbeskrivningar. Utbildningsinspektionen som ska granska skolans inre arbete kan också ses som inslag av typ 1- och typ 2-karaktär.

Delar av kursplanerna - främst de som gäller mål att uppnå och betygskriterier - kan tolkas i termer av standards.⁴² Likaså kan de ökande kraven på att samla in och publicera provresultat - som led i såväl nationell utvärderingsverksamhet som i lokal bedömning av den egna skolans och kommunens prestationer - ses som aktiviteter som närmast ingår i typ 4. Citatet ovan har således påtaglig relevans för de nationella provens roll i dagens svenska skolsystem. Å andra sidan anser Sivesind m.fl. att Sverige tycks vara på väg att ytterligare fördjupa sin inriktning mot typ 3, men då stöder de sig på tankegångar i utbildningsdepartementets skrivelse Samverkande styrning (Utbildningsdepartementet 2001).

Finland sägs för övrigt också tillhöra typ 3 men bedöms vara på väg mot både typ 1 (striktare föreskrifter) och typ 4 (ökad betoning på elevernas kunskapsmål). De norska läroplanerna karaktäriseras som typ 1 och 2, liksom de danska. I de nya danska läroplanerna finns dock även inslag av typ 4.

Svensk kunskapsbedömning under olika läroplaner

Nationella prov, nationella utvärderingar och internationella utvärderingar kan på olika sätt knytas till och motiveras av de läroplanstyper som beskrivits. Varje läroplanstyp genererar sina karaktäristiska instrument.

Lgr 68 och Lgy 70 kan närmast hänföras till läroplanstyp 1. Kunskapsuppdraget stod i centrum och kunskapsinnehållet var väl specificerat i läroplaner och supplement. Genom läroplanens tydligt föreskrivande roll förmedlades det kunskapsstoff som beslutats centralt. Intresset vid kunskapsprövningen var inte primärt riktat mot vad eleverna kunde, eftersom detta redan var föreskrivet. Instrumentens syfte var främst att på ett så korrekt sätt som möjligt särskilja elever och därmed ge ett tillförlitligt underlag för urval till högre utbildningar.

⁴² Se appendix.

Dnr 01-2003:2038

Det normrelaterade provsystemet med centrala prov för gymnasieskolan och standardprov för grundskolan var anpassat till denna typ av läroplan.

I och med Lgr 80 minskade centralstyrningen och ökade det lokala inflytandet. SIA-utredningen (SOU 1974:53) innebar att större uppmärksamhet riktades mot skolans inre arbete. Läroplanens innehållsangivelser blev mer allmänna och ställde större krav på lokal uttolkning. Betygssystemet var inte längre strikt normrelaterat med givna procentsatser för olika betyg, även om den femgradiga betygsskalan behölls. De normativa inslagen ökade också och kunskapsuppdraget var inte längre lika allenarådande.⁴³ Standardproven behölls dock även de, men med mildare föreskrifter om hur resultaten skulle användas. Någon stark betoning lades inte på nationella utvärderingar, men de internationella jämförande studierna hade ökat under 1970- och 80-talen och de svenska resultaten var inte alltid smickrande, i synnerhet inte i matematik som framstod som ett krisämne på 1980-talet.

För gymnasieskolan skedde inte någon motsvarande utveckling utan där innebar övergången från Lgy 70 till Lpf 94 ett språng direkt från typ 1 till typ 3 utan mellanlandning i typ 2.⁴⁴

I Lpo 94 och Lpf 94 är kunskapsuppdraget ytterligare nedtonat till förmån för ett mer normativt demokrati- och värdegrundsuppdrag. Decentraliseringstanken med lokalt friutrymme och deltagande målstyrning är grundläggande. I detta system passar inte det tidigare provsystemet med standardiserade och normrelaterade prov in. I synnerhet inte som kursplaner och betygskriterier, som tidigare nämnts, nu har karaktär av (performance) standards, dvs. verbalt formulerade utsagor om vilka kunskapskvaliteter som skulle uppnås för olika betyg, medan innehållet i ämnena (content standards) ska väljas av lärare och elever gemensamt. Någon föreskrift om det konkreta innehållet i undervisningen finns inte längre. Ett centraliserat nationellt prov med samma innehåll för alla passar alltså inte in i ett system enligt läroplanstyp 3. Här är det i stället lärarens egen kunskapsbedömning, anpassad till det valda stoffet, som är central. Ett nationellt instrument kan då inte bli föreskrivande utan ger endast exempel på hur kursmål och kriterier kan tillämpas på ett konkret stoff. Proven blir betygsstödjande genom förebildlighet.

Eftersom inga föreskrifter längre finns om undervisningens genomförande och innehåll (vilket under läroplanstyp 1 implicit antogs garantera kvaliteten genom att föreskrifterna följdes) måste utfallet nu kvalitets säkras på något sätt. Detta förmodades ske genom lokalt kvalitetsarbete via föreskrifter om skolplan, lokal arbetsplan och kvalitets-

⁴³ Naturligtvis har skolan alltid innehållit starka normativa och värdegrundsrelaterade inslag men de har inte varit tydligt framskrivna i styrdokumentet.

⁴⁴ Här kan Lvux 82 ses som en övergångsform mellan Lgy 70 och Lpf 94, där t.ex. kurssystemet introduceras.

Dnr 01-2003:2038

redovisning. Något särskilt utarbetat instrument för nationell utvärdering av måluppfyllelsen ingår inte i läroplanstyp 3 och förutsattes inte heller enligt Lpo 94.

Ganska snart visade det sig att systemet med lokal utvärdering och kvalitetssäkring inte fungerade. Kvalitetsredovisningarna var av bristfällig kvalitet eller fanns inte alls. Detta ledde 1998 till införandet av nationella kvalitetsgranskningar. Dessa gällde dock endast ett litet urval av kommuner och behovet av nationell redovisning av måluppfyllelsen ledde 2001 till beslut om att resultat från de nationella proven ska samlas in för samtliga elever i år 9. De ska också publiceras för allmän kännedom.

Samtidigt har beslut fattats om att betyg inte ska sättas på proven utan endast på de olika delarna i proven (gäller svenska/svenska som andraspråk och engelska). Hur dessa delprovresultat ska relateras till slutbetyg i ämnet anges inte. Det är lärarnas sak att avgöra vilken roll delprovsbetygen ska ha för den slutgiltiga bedömningen. Trots denna oklarhet förväntas delprovsbetygen fungera som nationella underlag för bedömning av måluppfyllelse. En intressant fråga är vilka slutsatser som kan dras av de publicerade resultaten. Vad ska de relateras till? När ska resultaten anses som bra? Hur ska resultaten förstås om slutbetyg och delprovsbetyg visar olika saker?

Man kan sammanfattningsvis säga att de nationella proven (främst i år 9) ska vara betygsstödjande genom exemplifiering samtidigt som de ska fungera som instrument för nationell utvärdering och bedömning av måluppfyllelsen, men då med avseende på vissa delmål.

Det lokalt baserade evalueringsystem som förväntas för ett system enligt läroplanstyp 3 tycks således inte vara tillräckligt ur ett nationellt perspektiv. För att stärka den nationella kontrollen och insynen har inspektion införts, vilket närmast för tankarna till läroplanstyp 1 eller 2, beroende på vad som ska inspekteras. De nationella provens betydelse har samtidigt förskjutits från att vara ett betygsstödjande instrument med fokus på lärarens yrkesutövning, mot att bli ett utvärderingsinstrument för olika administrativa nivåer. Från att vara "kunskapsbedömning *för* lärande" mot att bli "kunskapsbedömning *av* lärande". Från att främst ha en formativ roll mot att ha en summativ. Från att vara kunskapsbedömning enligt typ 3 mot att bli kunskapsbedömning enligt läroplanstyp 4.

De intryck man får vid en tillämpning av den norska typologin på svenska förhållanden är att det finns påtagliga spänningar i systemet. Betoningen i läroplanen ligger på normativa värdefrågor, decentralisering och lokala val av innehåll och metoder. Evalueringen av ett sådant system förutsätter kraftfulla lokala instrument för kvalitetsgranskning. Sådana finns uppenbarligen inte i den utsträckning staten önskar eftersom ett antal åtgärder vidtagits för att på nationell (och lokal) nivå tydliggöra utfallet av skolans arbete. De initiativ som tagits och den praktik som tar form tyder på att vi i praktiken (om än

Dnr 01-2003:2038

inte i teorin) är på väg att lämna läroplanstyp 3 för att likt Finland få en tudelning mot typ 1 och 2 - via den förstärkta inspektionen - och mot typ 4 genom den ökande betydelse som de nationella proven ges som indikatorer på måluppfyllelse, uttryckt i den form av standards som betyg representerar. Genom publiceringen av samtliga skolors prov- och betygsresultat bereds vägen för att ställa skolor till svars för resultaten ("accountability"), vilket är typiskt för standardbaserade skolsystem.

Om det enbart gällde att utvärdera skolsystemet på nationell nivå skulle det räcka med att låta utvärderingen gälla ett stickprov av elever, på samma sätt som för de internationella jämförande studierna (PISA, TIMSS, PIRLS osv.). Detta är visserligen internationella instrument för kunskapsutvärderingar baserade på standards (och därmed typ 4), men i ett nationellt perspektiv kan de mycket väl fungera som utvärderingar av kunskapsläget i det svenska skolsystemet med referens till andra länder och till tidigare resultat.

NU 2003 är en svensk urvalsstudie som, även om den inte utvecklats med detta syfte, kan ses som förenlig med ett system baserat på en läroplan av typ 3. Syftet är här mer att förstå variation än att redovisa aggregerade resultat. Det dilemma som kan uppstå i ett sådant typ 3 - system är vad en sådan kunskapsutvärdering ska ha för referenser när det varken finns nationella föreskrifter eller tydliga nationella standards. NU-studien bygger dessutom på instrument som utvecklades 1992 och alltså baseras på Lgr 80, vilket knappast gör bedömningen lättare. Man kan naturligtvis göra jämförelser mellan olika deltagande skolor och mellan resultat 1992 och 2003. Men ur ett uppföljningsperspektiv kan frågan bli vad eleverna ska prestera på nationell nivå för att man ska kunna säga att det är bra eller dåligt.

Betygsfrågan

Sivesind m.fl. snuddar endast vid betygsfrågan i sin beskrivning. Den kan dock inte förbigås i detta sammanhang och det kan vara intressant att diskutera betygs roll i olika läroplanstyper. Sverige har ju som tidigare nämnts rört sig från typ 1 via typ 2 till typ 3 (med inslag av typ 4).

Betygsfrågan har av tradition haft stor betydelse i den svenska skoldebatten och också varit föremål för flera utredningar. Det som är slående ur ett svenskt perspektiv när man granskar den internationella litteraturen inom det här aktuella området är hur liten uppmärksamhet som ägnas åt betyg. I litteraturen om standards är det målen för undervisningen, tydligheten i systemet, huvudmännens ansvar osv. som betonas. Om betyg sägs mycket lite. Mot detta kan ställas

Dnr 01-2003:2038

det stora intresse som i Sverige ägnats betygsfrågan genom åren med ett flertal utredningar som följt.⁴⁵

Perioden med Lgr 62, Lgr 68 (och Lgy 70), som alla kan hänföras till typ 1, var rätt oproblematiskt. Betygsskalan var fastställd med normalfördelningen som grund, vilket innebar att staten föreskrev procentsatser för hur stor andel av eleverna som skulle ha respektive betyg. Betygen skulle ange elevernas rangordning men så däremot inget om vad de kunde.

Likvärdigheten vilade på strikt standardiserade normrelaterade prov som styrde betygsnivå och spridning i respektive klass. Däremot betonades att de inte styrde den enskilde elevens betyg. Fördelningen inom klassen överläts till läraren men medelvärde för slutbetygen fick inte avvika från medelvärdet på provbetygen med mer än ett angivet värde. Om så var fallet krävdes en skriftlig motivering till avvikelser. Lärarens inflytande över betygssättningen var förhållandevis litet och den centrala styrningen påtaglig.

För de ämnen som inte hade nationella prov gällde att inte heller de slutbetygen fick avvika mer än ett angivet värde från värdet på ett angivet nationellt prov som ansågs närliggande.

I och med Lgr 80 och övergången till ett mer typ 2 - liknande system blev betygssystemet i grundskolan mindre strikt. Föreskriften om att en viss procentsats av eleverna skulle ha respektive betyg mildrades och i Lgr 80 angavs endast en ungefärlig fördelning.

Betygen skall ges i en femgradig skala, där medelbetyget är tre. Någon given procentsats för de olika betygsgraderna skall inte finnas. Normalt skall dock antalet fyror och tvåor i en klass vara fler än antalet femmor respektive ettor.

Medelbetyget tre gäller för samtliga elever. För den enskilda klassen kommer däremot medelbetygen att kunna avvika från riksgenomsnittet. Standardprov ställs till skolornas förfogande som riktningsgivare för betyget tre i svenska, matematik och engelska. (sid 39).

Här har alltså en uppluckring av det normrelaterade betygssystemet skett genom att betygsfördelningen inte längre styrs från nationell nivå annat än i mycket grova drag. Betygens fördelning har överlåtit

⁴⁵I Finland finns t.ex. inga nationella prov för betygssättning i grundskolan utan endast utvärderingar baserade på stickprov (där för övrigt inga enskilda skolors resultat publiceras). Vid ett PISA-möte våren 2002 ställdes en fråga hur man i Finland visste att man hade ett rättvist och likvärdigt betygssystem. Svaret var att det har man inte och uppenbarligen sågs inte detta som något att bekymra sig för. Det låga intresset för betyg betyder dock inte att rättvisefrågor är ointressanta i andra länder. I USA är t.ex. "fairness" en stor fråga, men då gäller det i huvudsak olika typer av prov, examensprov, intagningsprov, urvalsprov, certifieringsprov etc., och i mindre utsträckning betyg som sällan har den betydelse de har i Sverige. En annan iakttagelse är för övrigt att den i svensk retorik mycket frekventa användningen av begreppet kvalitet i olika kombinationer inte heller har någon tydlig motsvarighet i den internationella litteraturen om läroplaner, standards, utvärdering och kunskapsbedömning. Måhända är det en retorik som främst lämpar sig för ett typ 3-system.

Dnr 01-2003:2038

till lärarens professionella omdöme. Detta kan sägas vara i linje med ett av grunddragen i läroplanstyp 2, nämligen att i förlitan på kompetent personal mer betona grundläggande principer för skolans arbete och i mindre grad detaljföreskriva vad som ska göras, och hur det ska göras. De nationella provens roll blir reducerad i ett sådant system, men kan fortfarande fungera som riktningssgivare för nivån. Likvärdigheten i betygssättningen försvagas när tydliga riktlinjer finns för endast ett av de fem betygen, nämligen medelbetyget 3.

Den läroplanstyp som innebär de största problemen för en likvärdig betygssättning torde vara läroplanstyp 3. Här är det de politisk normerande värdena som är centrala medan kunskapsuppdraget endast beskrivs i allmänna termer. Den konkreta utformningen av innehåll och arbetsformer överläts till den individuella läraren (och eleverna). I den svenska versionen beskrivs målen på två nivåer, mål att sträva mot som ska ligga till grund för undervisningens uppläggning, och mål att uppnå som anger vad den enskilda eleven ska uppnå. De senare målen kräver dock, liksom betygskriterierna, lokal konkretisering innan de kan tillämpas.

Det finns minst två problem förknippade med betygssättning i ett system av typ 3, i varje fall i den svenska versionen. Det första gäller den grundläggande frågan vilken betydelse demokratimål och värdegrundsmål ska ha för betygssättningen. Det andra gäller likvärdigheten.

Värdegrundsmål och betyg

När det gäller värdegrundsmålen roll i betygssättningen kan problemet ställas på sin spets med t.ex. följande fråga. "Kan en elev med uttalat rasistiska uppfattningar få MVG i samhällskunskap?"

Det är ingen som helst tvekan om att rasistiska uppfattningar strider mot läroplanens värdegrund. Däremot är det svårt att finna att de strider mot mål att uppnå. Om rasistiska uppfattningar utifrån värdegrunden bedöms som betydelsefulla för betygssättningen, vilken betydelse ska de då ha? Kan eleven i stället för MVG få VG - eller G? Fortfarande står ju elevens uppfattningar i strid mot värdegrunden. Eller ska man gradera de rasistiska eller antidemokratiska uppfattningarna, så att om man bara är lite rasistisk eller lite antidemokratisk kan man få VG. Är man lite mer rasistisk eller antidemokratisk får man G, osv. Ska man med andra ord utforma betygskriterier för vilken grad av rasistiska och antidemokratiska uppfattningar som kan accepteras för olika betyg?

Eller ska man helt enkelt säga att om man överhuvudtaget uttrycker rasistiska uppfattningar kan man inte få något betyg i samhällskunskap? Men då blir nästa fråga: Kan en elev som uttrycker rasistiska värderingar få MVG i historia, eller i biologi, eller matematik, eller något annat ämne? Värdegrundsmålen gäller ju oberoende av ämne.

Dnr 01-2003:2038

Den mest rimliga slutsatsen måste vara att värdegrundsfrågorna lämnas utanför betygssättningen. Moralisk-etiska frågor lämpar sig inte för gradering eller bedömning i betygstermer. Frågorna måste givetvis behandlas i skolorna men om de läggs till grund för betygssättningen kommer elever snabbt att lära sig att uttrycka de politisk korrekta uppfattningarna, oberoende av vad de egentligen tycker. Men den typen av opportunism kan i sin tur knappast vara i linje med skolans värdegrundsmål.

Likvärdigheten

Det andra problemet avseende betygssättning i läroplanssystem av typ 3 gäller likvärdigheten i betygssättningen.⁴⁶ I detta system är den vaghet som framträder i typ 2 ytterligare förstärkt. Nationella prov ska användas men hur de ska användas anges inte. Betygssättningen är överlämnad till den enskilda läraren och till de lokala aktörerna. (När det gäller betygssättningen är dock det svenska systemet, som tidigare nämnts, inte renodlat enligt typ 3 utan innehåller betydande inslag av typ 4 i form av performance standards).

Likvärdigheten är dock problematisk. Skolverkets kvalitetsgranskning om betyg (Skolverket, 2000), Stockholm stads skolinspektörers rapporter och andra lokala rapporter tyder på att det finns betydande brister i likvärdigheten. Skolverkets panelstudier visar dock att många lärare själva tycker att de förstår målen och kriterierna och att de är tydliga. Det finns ingen anledning att betvivla att de som säger detta menar vad de säger. Problemet är snarare att det inte är säkert att olika lärare som tycker att mål och kriterier är tydliga sinsemellan är eniga om tolkningen. Så länge den egna tolkningen inte konfronteras med alternativa tolkningar framstår den lätt som tämligen oproblematisht.

Det samtalsbaserade system vi har, där tolkningen av mål och kriterier förutsätter ett ständigt pågående professionellt samtal, ställer stora krav på lärarna och skolorna, inte minst med avseende på den tid som måste avsättas. Ett system, som typ 3, som alltså ska förena kraven på lokal målutformning med nationell likvärdighet kräver förutsättningar i form av såväl arenor som tid för ett ständigt fortlöpande professionellt och kollektivt tolkningsarbete. Sådana förutsättningar finns uppenbarligen inte i tillräcklig utsträckning och likvärdigheten framstår därmed närmast som ett fjärran mål att sträva mot.

Men likvärdigheten är viktig. Betygens betydelse för t.ex. möjligheterna att välja gymnasieskola (eller högskoleutbildning om det gäller

⁴⁶ Med likvärdighet avses att likvärdiga kunskaper ska ge likvärdiga betyg. Begreppet likvärdighet har annars i den offentliga retoriken genomgått en betydelseförskjutning från att främst ha avsett de yttre generella förutsättningarna till att gälla den enskilda individens förutsättningar, dvs. från en strävan att alla skolor skulle erbjuda samma förutsättningar, till en strävan att alla elever ska erbjudas det stöd som behövs för att de som individer ska uppnå målen.

Dnr 01-2003:2038

gymnasiet) gör givetvis att rättvis- och likvärdighetsfrågor är ständigt aktuella bland lärare, elever, föräldrar m.fl. De nationella proven förväntas av många fungera som en normerande faktor - inte minst som en följd av att de flesta vuxna själva gått i en skola med normerande prov - men som tidigare visats är provens betygsstödjande roll inom ramen för ett system av typ 3 i huvudsak att fungera som förebild och exempel för lärarens egna klassrumsbaserade bedömningar.

När det gäller läroplanstyp 4 så finns som redan nämnts inslag i det svenska systemet som kan hänföras till denna kategori. Mål att uppnå och betygskriterierna kan karaktäriseras som standards. Syftet med dessa svenska standards är dock inte att ligga till grund för utvärderingen av skolsystemet (det har i varje fall inte varit det primära syftet) utan att ligga till grund för lärarnas översättning av de nationella målen och betygskriterierna till lokala konkretiserade motsvarigheter. Det är i denna översättningsprocess som det är svårt att uppnå likvärdighet.

Dnr 01-2003:2038

Utredningsgruppens bedömning

Även om läroplanstyp 4 framför allt syftar till att formulera standards för utvärdering av det nationella utbildningssystemet så är det givetvis möjligt att utforma tydliga standards som också kan ligga till grund för betygssättning av enskilda elever. Det svenska systemet är dock inte tydligt i detta avseende eftersom det skulle stå i strid med principen om den lokala målutformningen.

I dagens spänningsfyllda svenska system som i teoretiskt hänseende i huvudsak hör hemma i läroplanstyp 3 men som har starka rötter i läroplanstyp 1 (inte minst hos föräldrar och andra vuxna) och som på central nivå rör sig mot läroplanstyp 4, är det förståeligt om olika aktörer har olika uppfattningar om det nationella provsystemets utformning och syfte.

I sammanfattning skulle ovanstående resonemang innebära

att proven av många avnämare förväntas genomföras under de standardiserade former som gällde för de tidigare normrelaterade proven, och att de ska ge inbördes jämförbara provresultat (typ 1),

att proven förväntas tjäna som förebilder för lärarnas egna klassrumsbedömningar medan de aktuella provresultaten kan vara missvisande på grund av annat stoffval än det i provet ingående (typ 3),

att provresultat förväntas kunna aggregeras och publiceras på såväl nationell som lokal nivå för att tjäna som underlag för att bedöma skolornas måluppfyllelse och de ansvarigas insatser (typ 4).

Dessa förväntningar är inbördes svårförenliga inom ramen för ett enda instrument och provens roll blir därmed oklar. Det är osäkert om de under rådande regelsystem fungerar tillfredställande i något av de tre avseendena.

Den rimliga slutsatsen av ovanstående resonemang blir att den nationellt organiserade kunskapsbedömningens syften måste preciseras. Först när detta är gjort kan provsystemet utformas på ett konsekvent och ändamålsenligt sätt.

Dnr 01-2003:2038

5. Nationella prov i flera ämnen och för flera åldrar?

En del i det regeringsuppdrag som ligger till grund för utredningen är att undersöka förutsättningar för att vidga provsystemet till flera åldersgrupper i grundskolan och till flera ämnen i samtliga skolformer.

Provsystemets tolkas i det här sammanhanget i vidgad mening, dvs. det tilldelas inte endast den funktion som de nuvarande nationella proven har utan innefattar också nationella och internationella kunskapsutvärderingar.

I det här kapitlet redovisas i första delen lärares och elevers syn på en vidgning av provsystemet till fler ämnen och åldrar. Redovisningen bygger på resultat från den effektstudie av provsystemet som tidigare nämnts, främst i kapitel 2, och från en hearing som genomförts med lärare och elever.

I den senare delen av kapitlet förs en del resonemang om en vidgning av provuppdraget till flera ämnen genom närmare samordning mellan olika typer av kunskapsutvärderingar och nationella prov.

Prov i flera ämnen och för flera åldrar?

Frågor om prov i fler ämnen och för flera åldrar ingår i den effektstudie som tidigare relaterats till i kapitel 2. Frågan har också diskuterats vid en hearing till vilken representanter för olika skolämnen inbjudits tillsammans med elevrepresentanter.

Vill lärare ha prov i flera ämnen?

Enkätstudie i grundskolan

Resultaten från den tidigare omnämnda enkätstudien till lärare och rektorer i grundskolan som genomfördes våren 2003 är tydliga. Man är nöjda med de nuvarande proven och diagnostiska materialen men visar inget nämnvärt intresse för prov i andra ämnen eller för andra åldrar. Tabell 9 visar en sammanställning av svaren.

Dnr 01-2003:2038

Tabell 9. Lärares och rektorers svar på frågan "Hur viktigt är det att nationella prov och diagnosmaterial finns?" Svarsfördelning i procent.

Hur viktigt är det att det finns...?	Mycket viktigt		Ganska viktigt		Ganska oviktigt		Inte alls viktigt	
	Lär	Rekt	Lär	Rekt	Lär	Rekt	Lär	Rekt
Ämnesprov år 9	54	70	41	27	5	2	1	1
Ämnesprov år 5	45	61	45	34	8	4	2	1
Diagnosmaterial F-5	28	41	53	47	15	10	4	2
Diagnosmaterial 6-9	27	37	52	50	17	11	4	2
Nationella prov för andra år	9	7	20	17	44	51	28	25
Nationella prov i andra ämnen	7	7	20	21	44	44	29	29
Diagnostiska material i andra ämnen	9	11	32	29	37	39	23	21

Som framgår av tabellen tycker över 90 procent av rektorer och lärare i grundskolan att nuvarande prov är mycket eller ganska viktiga. Även stödet för de diagnostiska materialen är stort.

När det däremot gäller intresset för prov för andra åldrar eller andra ämnen är intresset mycket begränsat. Mellan 25 och 30 procent av lärarna tycker att det inte alls är viktigt och nästan hälften att det är ganska oviktigt. Mindre än 10 procent tycker att sådana prov skulle vara mycket viktiga.

Man kan alltså konstatera att det i grundskolan finns ett starkt stöd för de nuvarande nationella proven och diagnostiska materialen, men att det inte finns tillnärmelsevis samma intresse för prov och diagnostiska material i andra ämnen eller för andra åldrar.

För gymnasieskolans del finns inte något motsvarande underlag. Provinstitutionerna bifogar frågeformulär till lärarna men där ingår inte någon fråga om hur viktiga man tycker att de nationella proven är.

Hearingen

Den hearing som genomfördes inom ramen för uppdraget innefattade cirka 20 lärare i olika ämnen samt två representanter för elevorganisationer. Denna hearing kan givetvis inte ses som representativ men gav goda exempel på den variation av uppfattningar som finns.

Även i det här sammanhanget kunde man allmänt konstatera att de ämnen som idag har nationella prov ser dessa som en stor tillgång. Mest uttalat var detta av representanten för engelska och moderna språk. Matematik representanterna var något mindre entusiastiska, någon ansåg att proven i alltför stor grad styrde arbetsformerna och bidrog till att t.ex. laborativ matematik fick stå tillbaka. Det förekom också visst ifrågasättande av att det behövs så många prov i gymnasieskolan.

Även de representanter för samhällsorienterande ämnen som deltog uttryckte intresse för någon form av nationella utvärderingsinstrument eller nationella prov. Däremot var representanten för ekono-

Dnr 01-2003:2038

miska ämnen negativ och ansåg att prov skulle komma att styra undervisningen alltför mycket.

Inom den grupp som representerade olika ”praktisk-estetiska” ämnen var uppfattningarna delade. Musik- och bild- representanter såg fördelar i ett nationellt prov genom att de kunde bidra till att ge ämnena bättre status och ställning i skolan. Representanten för slöjdämnet däremot såg inget behov av något nationellt prov och representanten för hem och konsumentkunskaps ställde sig mycket tveksam och uttryckte en stark oro för att ämnet skulle inskränkas genom att resultat mycket mer än process skulle komma i fokus.

Flera representanter som förordade prov framhöll den ökande andelen obehöriga lärare som ett skäl. För denna kategori av lärare skulle nationella prov vara ett särskilt viktigt stöd i bedömningen av elevernas måluppfyllelse.

Det framfördes också oro för att en alltför omfattande provverksamhet blir för tidskrävande, både för att eleverna förlorar undervisningstid och för att lärarnas rättningsbörda ökar.

Några representanter tyckte att det var viktigare med tydligare mål och kriterier än fler prov. Om man ska utgå från proven blir de alltför styrande och hämmar den pedagogiska utvecklingen. Det fanns dock andra röster som såg nationella prov som viktiga instrument för att utveckla det pedagogisk-didaktiska tänkandet.

Tanken på att i stället för prov tillhandahålla kommentarmaterial med provuppgifter och kommenterade autentiska elevlösningar väckte rätt stort gensvar. Detta kan sägas vara ett av de nuvarande provsystemets syften, men sekretessreglerna gör att prov, elevlösningar och bedömningsanvisningar inte kan publiceras och diskuteras, och därmed endast i begränsad omfattning kan ses som ett sådant kommentarmaterial.

Det fanns också önskemål om bedömningsmaterial i t.ex. teckenspråk.

Vad anser eleverna om prov i flera ämnen?

Effektstudien

Någon enkät har inte riktats till eleverna på grundskolan när det gäller prov och diagnosmaterial. Däremot har intervjuer gjorts med elever vid några skolor.

Den allmänna uppfattningen bland de elever som intervjuades var att man borde ha prov i alla ämnen för att mäta målen men de ansåg att det ändå var kärnämnen som var viktigast, där skulle proven vara kvar. De tyckte samfällt att det var viktigt med nationella prov.

Dnr 01-2003:2038

Hearingen

De två elevrepresentanter som deltog i hearingen var uttalat negativa till de nuvarande nationella proven. De ansåg att proven användes på mycket olika sätt av olika skolor och av olika lärare. Eleverna visste ofta inte hur proven skulle se ut och vad de skulle omfatta, eller hur provresultaten skulle användas. Proven skapar stress hos eleverna och det är önskvärt att proven avdramatiseras. I princip var de dock inte mot prov men ansåg att eleverna skulle ha ett inflytande över om de skulle användas eller ej. Ett provbankssystem med större flexibilitet än dagens nationella prov förespråkades. De nuvarande proven upplevs ofta som examensprov för att de tagits fram på central nivå. Speciellt proven i matematik upplevs som "laddade" och som bidragande orsak till att matematik ses som ett teoretiskt pluggämne.

Eleverna ansåg vidare att det arbetssätt man tillämpade ofta inte stämde med det arbetssätt som proven förutsatte. Denna uppfattning kan sägas ha viss stöd från lärarhåll i den tidigare nämnda enkätstudien, där lärare i högre grad anser att de nationella proven påverkar arbetsformer än innehåll i undervisningen.

Eleverna talade för lokala utvärderingar med högre grad av kvalitativa inslag än vad som är möjligt med centralt konstruerade nationella prov. Man kan säga att eleverna förespråkade ett kvalitetsgranskningsystem med stark lokal prägel, vilket kan sägas var väl i samklang med de ursprungliga intentionerna för ansvarsfördelningen inom skolsystemet.

Eleverna framförde också att nationella prov lägger fokus på den enskilda eleven i stället för på skolan som arbets- och studiemiljö, dvs. de leder till att utvärderingen får ett individperspektiv i stället för ett skolperspektiv.

Något stöd för tanken att de nationella proven kan ses som ett sätt att säkra elevernas rättsäkerhet när det gäller betygssättningen uttryckte inte de deltagande elevrepresentanterna.

Slutsatser

För grundskolans del är bilden rätt klar. Såväl lärare som rektorer och elever tycker det är viktigt att det finns nationella prov och diagnostiska material i de ämnen och årskurser som är aktuella idag. Nästan lika tydligt är det att rektorer och lärare *inte* önskar prov i nya ämnen eller för nya åldrar. Man bör dock ha i åtanke att enkäten riktar sig till lärare som undervisar i de nuvarande provämnena svenska/svenska som andraspråk, engelska och matematik. Detta kan ha betydelse för synen på prov i andra ämnen. För rektorens del finns dock ingen sådan snedhet utan de kan representera alla ämnen.

För elevernas del finns inget tydligt underlag när det gäller synen på prov i nya ämnen. Vissa intervjudata kan dock tyda på att de är mer

Dnr 01-2003:2038

positivt inställda än lärare och rektorer. Likaså är det okänt vad föräldrar anser i frågan.

För gymnasieskolans del är bilden mer oklar. Det saknas ett tillförlitligt empiriskt underlag och slutsatserna vilar i huvudsak på information från en hearing och från olika former av mer eller mindre anekdotisk information.

Det förefaller dock som om bilden är mer splittrad. Vissa företrädare för nuvarande provämnen uttryckte starkt stöd för den rådande ordningen och befarade att en ändring skulle skapa stort motstånd hos lärarna, andra var mer tveksamma till provens positiva konsekvenser.

Av de ämnen som inte har prov i dag fanns också blandade uppfattningar. Vissa vill ha prov andra såg konsekvenserna som mer negativa, genom att proven blir mer styrande än stödjande.

De elevrepresentanter som deltog i hearingen var uttalat negativa till de nationella proven av flera skäl. I vilken utsträckning deras uppfattningar återspeglar uppfattningar hos majoriteten av elever är svårt att veta. De grundskolelever som intervjuats i samband med studien av provsystemet i grundskolan är till övervägande del positiva till de nationella proven.

Man måste naturligtvis respektera lärares och rektorers uppfattningar om att det är viktigt med nuvarande prov men oviktig med prov i andra ämne. Det är dock svårt att veta vad man tänkt sig för typ av prov i andra ämnen när man besvarat enkäten. Med tanke på att lärarna nu ofta klagat över den stora arbetsinsats som krävs för att genomföra proven är det förståeligt om de ryggat för ytterligare prov.

Skolverket får tidvis via brev och telefonsamtal vissa propåer om prov i andra ämnen än de nuvarande. De grupper som under senare tid framför allt kontaktat Skolverket är organisationer med anknytning till kemiämnet. De uttrycker en allmän oro för ämnets utveckling och ser provbanksprov i kemi för gymnasieskolan som viktiga för att stärka ämnet och dess ställning.

Breddning med hjälp av andra instrument än nationella prov

Att följa upp och utvärdera kunskaper sker inte bara med hjälp av nationella prov, det görs också i form av internationella jämförande studier och nationella utvärderingsstudier. Exempel på de förstnämnda är Pisa, Timss och Pirls, exempel på det senare NU-03.

Internationella studier

Pisa är en Oecdstudie av 15-åringars kunskaper i läsförståelse, matematik och naturvetenskap. Timss är en IEA-studie av matematik och science för elever i klass åtta (13-14 åringar). Även Pirls är en IEA-studie av läsförståelse för elever i årskurs 4 (9-10 åringar).

Dnr 01-2003:2038

Pisa genomförs för närvarande i treårsintervall. Datainsamlingar är gjorda 2000 och 2003 och en tredje planeras för 2006. Även övriga internationella studier planeras att genomföras med bestämda tidsintervall, inte minst för att det finns en önskan om att kunna följa kunskapsutvecklingen över tid.

De internationella studierna har sin främsta styrka i att de ger referenspunkter för jämförelser av kunskapsläget i olika länder. Det är också genom att många länder samverkar möjligt att till en förhållandevis låg kostnad få tillgång till den mest avancerade expertkompetens som finns att tillgå inom kunskapsutvärdering.

En annan fördel med att många länder samverkar är att resultaten ofta kan följas upp genom olika tematiska studier på ett sätt som enskilda länder har svårt att genomföra på egen hand. Sådana rapporter kan sedan översättas och modifieras för att nå en bredare publik än den som nås med engelskspråkiga rapporter.

En nackdel med internationella studier är naturligtvis att de i viss mån blir kompromisser av de ingående ländernas önskemål. De aktuella instrumenten är därmed inte anpassade till de enskilda ländernas förutsättningar, vilket gör att de inte blir optimala ur ett nationellt perspektiv.

De internationella studierna överlappar delvis de nationella proven men tillhandahåller också data i ämnen där det för närvarande inte finns nationella prov, främst gäller det naturvetenskapliga ämnen som provas i både Pisa och Timss.

Nationell kunskapsutvärdering

Nationella studier av typen NU har fördelen att kunna anpassas till svenska förutsättningar. Nackdelen är då snarast att jämförelsepunkter saknas. Det kan vara svårt att avgöra hur resultaten bör bedömas. Det är inte heller möjligt att med begränsade resurser anlägga en lika rigorös teknisk hantering som för de internationella studierna.

Den nu pågående nationella utvärderingen omfattar alla ämnen i grundskolan utom moderna språk och teknik samt ett antal enkäter till elever, rektorer lärare och föräldrar. Detta gör att väldigt stora datamängder insamlas. Fördelen med den rika tillgången på data är givetvis att det finns många analysmöjligheter. Nackdelen är att det är svårt att hinna göra alla dessa analyser medan data fortfarande har sin aktualitet.

Nationella utvärderingar tillhandahåller dock data för ett stort antal ämnen som idag saknar nationella prov, främst gäller det samhällsvetenskapliga och naturvetenskapliga ämnen, men även ämnen som bild, musik, slöjd samt hem- och konsumentkunskap. På det sättet kompletterar de nationella utvärderingarna de nationella proven.

Dnr 01-2003:2038

Sammanfattning

Det finns mycket ringa stöd bland rektorer och lärare i grundskolan för att vidga det nuvarande provsystemet till fler ämnen och åldrar. Detta gäller då i huvudsak uppfattningar hos lärare som har de nuvarande provämnena och hos rektorer. Vad t.ex. lärare i praktiskt-estetiska ämnen och samhällsvetenskapliga ämnen anser är mer oklart. Man kan dock inte i det här läget säga att det finns tydliga önskemål från någon lärargrupp eller intressegrupp för ytterligare nationella prov i grundskolan. Utredningsgruppen bedömer därför inte att det finns skäl att vidga det nuvarande provsystemet till att omfatta nationella prov av den nu aktuella slaget till fler ämnen och åldrar. Om man ser de nationella proven som ett instrument för likvärdig betygssättning är detta naturligtvis en svaghet. Likvärdigheten är inte viktig enbart i de nuvarande provämnena även om de är behörighetsgivande för gymnasieskolan. För meritvärderingen finns det i grundskolan ytterligare 13 betyg där det inte finns något nationellt prov.

För gymnasieskolans del finns det via provbanken tillgång till prov i ett antal ämnen utöver de tre obligatoriska provämnena. Här tycks också prov i flera ämnen vara mer önskade än i grundskolan även om bilden inte är entydig. Ingen ämnesförening eller intresseförening har framfört tydliga önskemål om ytterligare nationella prov med undantag för kemilärarnas olika föreningar och andra kemiintressenter.

Ur ett uppföljningsperspektiv är det givetvis en svaghet att endast tre ämnen ger underlag. I det avseendet finns dock internationella studier och studier av typen NU-03 att tillgå. Man kan också tänka sig en form av studier som utgör en hybridform av sällan återkommande heltäckande studier (typ NU) och mer kontinuerliga studier med få ämnen (typ Pisa). Till exempel kan man ha ett system med stickprovsvisa uppföljningsstudier i vissa ämnen med vissa års mellanrum. Första året några ämnen, andra året några andra, tredje året ytterligare några andra osv. Man kan välja en cykel på tre, fyra, fem år eller vad som bedöms lämpligt. Ett sådant system skulle fungera utmärkt för nationell uppföljning.

Materialet skulle vara sekretessbelagt men kunna användas lokalt av de skolor eller kommuner som så önskar. Man kan också tänka sig att materialet används som underlag vid Skolverkets inspektioner genom att skolor som ska inspekteras uppmanas att genomföra några av kunskapsutvärderingarna som ingår. Något större tryck på sekretessen torde det inte bli eftersom detta material inte är kopplat till betygssättning. Sekretessen bidrar också till att trendmätningar över tid med lätthet kan utföras. Denna typ av studie lämpar sig också väl för roteerade häften, vilket innebär att ett stort kunskapsområde kan täckas in trots att varje elev endast arbetar med ett relativt begränsat material. På detta sätt blir belastningen på deltagande skolor betydligt mindre än om alla elever ska delta i alla prov.

Dnr 01-2003:2038

Utredningsgruppens bedömning

Inga nya nationella prov bör införas i grundskolan.

Vid en eventuell utvidgning av prov i gymnasieskolan bör kemiämnet uppmärksammas.

Möjligheten och kostnaden för att genomföra rullande sticksprovsbaserade utvärderings- och uppföljningsstudier i olika ämnen eller kunskapsområden bör utredas.

6. Några avslutande synpunkter, sammanfattning och förslag till ställningstaganden

Så vilka blir då slutsatserna? Utredningen har vridit och vänt på det nationella provsystemet för att granska det ur olika perspektiv. Är provsystemet bra som konkretisering av de nationella kursmålen och betygskriterierna? Bidrar provsystemet till en rättvis och likvärdig betygssättning? Är provsystemet ett bra stöd för läraren i hans eller hennes förståelse av elevernas lärande? Hänger provsystemets olika delar samman? Är budskapen klara och tydliga? Är provsystemet med andra ord av hög kvalitet?

Signalier

Det finns vissa svårförenliga signaler. Proven i år 9 och kursproven i kärnämnen och de avslutande karaktärsämnena är obligatoriska för lärarna att använda. Det är ett tydligt budskap. Men hur resultaten ska användas är helt och hållet en fråga för den enskilda läraren, om man inte lokalt bestämt något annat. Signalen blir: du måste göra proven men vad du gör med resultaten är din ensak. Det är inte ett tydligt budskap.

Proven ska konkretisera kursplaner och betygskriterier. Det finns inga skäl att tro att de inte gör det. Proven får genomgående goda betyg av såväl lärare som av externa granskare. De har väl genomarbetade uppgifter och utförliga och nyanserade bedömningsanvisningar. De är föredömliga kommentarmaterial av hög kvalitet. Men de hemligstämplas i tio år.

Kvalitet

Provsystemets kvalitet kan som nämndes i utredningens inledning betraktas ur olika perspektiv. Ett är att bedöma det på det sätt som gjorts i de olika kapitlen i utredningen. Ett annat är att se själva instrumenten för sig och det system i vilka de ingår för sig.

Med utgångspunkt i den senare uppdelningen är utredningsgruppens generella bedömning att själva proven genomgående är av hög kvalitet. Det gäller deras kvalitet som uttolkare av kursplaner och betygskriterier. Det gäller deras kvalitet som stöd för läraren i klassrumsarbetet. Även de diagnostiska materialen bedöms fungera väl.

Lärare, rektorer och elever förefaller därtill nöjda med proven. Av de enkäter som genomförts finns det anledning tro att lärarna tycker att proven på ett rätt tydligt sätt klargör vilka kunskapskvaliteter som olika betyg representerar och att proven i huvudsak bekräftar lärarnas egna bedömningar av elevernas kunskande. Och så är det ju om man ser på statistiken i kapitel 3, cirka 70-80 procent av eleverna har

Dnr 01-2003:2038

samma slutbetyg och provbetyg. De flesta verkar med andra ord nöjda med de nationella proven.

Därmed borde väl allt vara frid och fröjd, om det inte vore för tvivlen om provens roll för en rättvis och likvärdig betygssättning. Skolverkets kvalitetsgranskningar och olika tillsynsuppdrag, liksom inspektörsrapporter från t.ex. Stockholms stad har gett anledning till tvivel på hur likvärdiga betygen egentligen är. Kapitel 3 i utredningen reser också ett antal frågor i samma riktning. Är således betygssystemet likvärdigt och rättvist?

Utredningsgruppen ser detta som den svåraste frågan och vill därför diskutera provens och provsystemets roll i detta sammanhang.

Likvärdighet och rättvisa

Det svar utredningsgruppen kommer fram till när det gäller likvärdigheten och rättvisan i betygssystemet är att något säkert svar inte kan ges, men att det finns mer som talar för än emot att det finns brister i likvärdigheten. Slutsatsen blir att vi har ett betygssystem där likvärdigheten och rättvisan kan ifrågasättas.

Hur kan då lärarna och rektorerna vara nöjda med proven om mycket talar för att de inte resulterar i att betygssättningen blir rättvis och likvärdig? Resultaten i kapitel 2 visar att lärarna anser att proven fyller sina syften, däribland att stödja en likvärdig och rättvis betygssättning. Inte heller i övrigt framförs några tydliga signaler från lärarhåll att man inte är tillfreds med den egna betygssättningen. Till detta kan finnas flera förklaringar. En kan vara att man tycker sig ha en god förståelse för hur betygskriterierna ska tolkas och vilka kunskaper eller kunskapskvaliteter olika betyg motsvarar. Det tycker man kanske på skola A, skola B och skola C och alla är nöjda. Kanske har man inom respektive skola kommit fram till någon form av konsensus i sina tolkningar, eller kanske varje lärare sköter sin egen betygssättning utan inblandning från kollegor. Allt fungerar väl så länge man inte behöver konfronteras med de andra lärarnas eller skolornas tolkningar, dvs. så länge man inte samråder med varandra om tolkningen av betygskriterierna. Sådant görs i liten utsträckning enligt t.ex. Skolverkets kvalitetsgranskningar, och därmed framträder inte tolkningsskillnaderna. Frid råder.

Men vad händer då när de nationella proven dyker upp? I många fall visar de enligt kapitel 3 andra resultat än de som den egna bedömningen gett. Vilka resultat väger då tyngst, provbetyg eller egna bedömningar? Uppenbarligen finns det skillnader mellan ämnena. I engelska varierar bedömningen mellan olika skolor, ibland blir slutbetyget högre än provbetyget, ibland lägre. I matematik är bilden mer entydig även om variationen även i det fallet är betydande. Men hur den enskilde läraren diskuterar när han eller hon ska ta ställning till elevernas provresultat vet vi inte. Vi vet inte heller hur, eller i vilken

Dnr 01-2003:2038

utsträckning, frågan diskuteras på rektorsnivå eller kommunnivå. Detta är en brist.

I matematik tycks man lita mer på den egna bedömningen. Eller gör man inte det? Finns det andra förklaringar till skillnaden mellan provbetyg och slutbetyg än att vissa lärare inte bryr sig om provbetygen? Som visats i kapitel 3 finns det ett antal mer eller mindre rimliga förklaringar till att betygen kan vara olika och återigen blir svaret att sådana avvikelser kan inte bedömas om man inte känner till de lokala förhållandena. Om man tror på detta är det inte rimligt i ett skolsystem med mål- och resultatstyrning av svensk modell att postulera att endast en viss avvikelse är acceptabel. Relationen mellan provbetyg och slutbetyg måste bedömas lokalt.

Även ur ett uppföljningsperspektiv är ett mekaniskt insamlande av provbetyg och slutbetyg föga meningsfullt om inte relationen mellan de olika betygen kan kommenteras och diskuteras. Vilket av betygen ska ses som det som bäst beskriver måluppfyllelsen? Om provbetyg väljs kommer vissa skolor att framstå i bättre dager, om i stället slutbetyg väljs blir det andra skolor som verkar mer framgångsrika. Återigen blir slutsatsen att redovisningen av betyg måste kompletteras med annan information och detta kan endast göras lokalt.

Att likvärdiga och rättvisa betyg är viktiga torde ingen förneka. I synnerhet när möjligheten till attraktiva högre utbildningar kan hänga på en decimal i betygsmedelvärdet är det naturligtvis nödvändigt att betygen är så rättvisa och likvärdiga som möjligt.

Likvärdighet

Men vad innebär det då att betyg ska vara likvärdiga och att de ska vara rättvisa? Likvärdighetsbegreppet har i skolan utvecklats från att främst ha syftat till att ge olika grupper samma förutsättningar att ta del av utbildning, enhetsskola, grundskolans utbyggnad, en gymnasieskola för alla, komvux osv. kan ses som led i en sådan strävan efter likvärdighet på gruppnivå. Man kan se en sådan likvärdighet som ett inslag i den sociala ingenjörskonsten.

Efterhand har dock likvärdighetsbegreppet alltmer kommit att bli en del i det personliga projektet. Idag handlar inte likvärdighet bara om att alla ska ha samma förutsättningar utan snarare om att var och en ska ges det stöd eller förutsättningar han eller hon behöver. En likvärdighet som innebär åt var och en efter behov.

Denna begreppsförskjutning är kanske svår att tillämpa på betyg och betygssättning. De flesta torde med likvärdig betygssättning mena att två personer med samma betyg ska ha samma kunskaper, vilket förefaller vara en rimlig tolkning. Eller gör det inte det? Vad innebär "samma" kunskaper? Betyder det att två elever med samma betyg ska kunna precis samma saker (inom vissa rimliga gränser)? Eller kan de kunna olika saker som har lika värde, det vill säga ha kunska-

Dnr 01-2003:2038

per som är likvärdiga men olika, som uttrycker samma kunskapskvaliteter men om olika saker? Och kan då samma prov ge båda likvärdiga förutsättningar att visa sina olikformiga men lika kunskapskvaliteter?

Rättvisa

Vad innebär då en rättvis betygssättning? En närliggande tolkning är att det handlar om att en elevs betyg verkligen ska spegla elevens kunskaper och kunskapskvaliteter så som de uttrycks i skolans styrdokument. Här uppstår då två tänkbara svårigheter, dels att tolka styrdokumentet på ett "likvärdigt" sätt, dels att bedöma eleven på ett "rättvist" sätt. Problemen med tolkningen av styrdokument har behandlats i andra sammanhang i utredningen och behöver inte upprepas här. Men den rättvisa bedömningen kan vara värd några ord.

Att bedöma någons kunskaper är naturligtvis en mycket grannlig uppgift, att därtill göra det på ett rättvist sätt är än svårare. En lärare kan rimligtvis inte bedöma annat än en elevs utsagor, muntliga eller skriftliga, eller elevens handlingar. Ett traditionellt sätt att försöka skapa rättvisa är att på olika sätt standardisera bedömningssituationen. Till exempel genom att föreskriva att alla ska genomföra vissa uppgifter under vissa givna förhållanden. I provsammanhang innebär det att alla gör samma prov, att alla får samma tid på sig, att inga hjälpmedel får användas (annat än vissa föreskrivna), att samtal med andra provdeltagare är otillåtet. Och givetvis är kontakt med världen utanför provsalen förbjuden.

I detta slutna rum ska eleverna genomföra ett antal uppgifter eller lösa ett antal problem. Det vill säga villkoren för att genomföra uppgiften avviker fullständigt på det sätt på vilket man löser problem i den verkliga världen (dvs. världen utanför provsalen). Där är det naturligtvis en värdefull kompetens att kunna utnyttja hjälpmedel och att ha ett nätverk av kunniga kontakter som kan rådfrågas. Är det då rättvist att pröva måluppfyllelsen under sådana standardiserade villkor? Ger resultaten av denna standardiserade och "likvärdiga" provningsprocess en rättvis bild av vars och ens måluppfyllelse?

Slutsatser

Utredningsgruppen accepterar att prov behöver finnas. Stora grupper behöver ibland provas i ett eller annat avseende och av praktiska skäl kan det vara nödvändigt att genomföra prov. Det kan också vara viktigt att använda prov för att få en generell bild av kunskapsläget för olika grupper. Men utredningsgruppen anser inte att prov av det slag som beskrivits ovan är bra verktyg för att avgöra måluppfyllelse hos och betyg för enskilda elever, inte som målen är formulerade i styrdokumentet. Det betyder inte att proven i sig inte är bra. Det betyder att den artificiella miljö i vilka proven genomförs inte ger alla individer samma förutsättningar att visa sitt kunnande, det är t.ex. välbe-

Dnr 01-2003:2038

kant att pojkar i allmänhet lyckas bättre än flickor på denna typ av prov. Det betyder också att de resultat som erhålls inte fångar in väsentliga egenskaper som förmåga att på ett kreativt sätt utnyttja olika hjälpmedel (och inte bara den tillåtna formelsamlingen eller lexikonet), eller att med hjälp av andra människor söka sig fram till svar.

Vad som är likvärdigt och rättvist låter sig således inte på något enkelt sätt fångas in genom att alla ges samma förutsättningar och genom att alla genomför ett prov under exakt lika villkor. Att föreskriva att ett standardiserat provresultat är det mest likvärdiga och rättvisa sättet att bedöma och betygssätta en människa är således enligt utredningsgruppens bedömning en alltför förenklad syn på vad rättvis och likvärdig betygssättning innebär.

Så vad är då en mer rättvis och likvärdig betygssättning? Utredningsgruppen anser att det är ett nödvändigt villkor att bedömningen av en elevs kunskaper görs av en kvalificerad bedömare som till sin hjälp har olika former av underlag, varav naturligtvis skriftliga och muntliga utsagor producerade under standardiserade villkor kan utgöra ett. Men det sistnämnda får inte vara det enda. Om olika elever med samma kunskaper ska bedömas på samma sätt av olika lärare måste det också finnas någonting tydligt att förhålla sig till vid betygssättningen. Även om läraren har en mycket god bild av vad en elev kan så blir betygssättningen otillförlitlig om läraren inte förstår vad som avses i de nationella styrdokumentet.

Det finns alltså två poler i bedömningen. Läraren måste få en så god bild som möjligt av elevens kunskaper och läraren måste förstå innebörden i det som elevens kunskaper ska relateras till. De nationella proven kan vara viktiga verktyg i båda avseendena, men de kan enligt utredningsgruppens bedömning inte bära hela bördan. Elevens resultat på provet måste kompletteras med lärarens kvalificerade bedömning. Och provkonstruktörernas tolkning av de nationella styrdokumentet måste kunna granskas, bedömas och analyseras utifrån rimligt tydliga nationella styrdokument. Enligt utredningsgruppens bedömning måste de nuvarande styrdokumentet kompletteras med exemplifierande kommentarmaterial som kan ligga till grund för såväl konstruktionen av nationella prov som lärarens egna bedömningar i det löpande arbetet.

Ett enda instrument kan inte uppfylla alla syften som läggs på provsystemet. Det är nödvändigt att inom systemets ram använda olika instrument för olika syften. Endast på det sättet kan det nationella provsystemet bli av hög kvalitet när det gäller de olika syften det har: att stödja likvärdig och rättvis betygssättning, att tydliggöra styrdokumentens kunskapssyn, att ge underlag för uppföljning av resultat på olika nivåer, att stödja lärarna så att elevernas lärande blir det bästa möjliga.

Att få likvärdighet och rättvisa i betygssättningen är således ingen enkel process. Samtidigt används betyg för olika former av urval.

Dnr 01-2003:2038

Ambitionen måste då vara att betygen ska vara så jämförbara som möjligt. I det nuvarande systemet baseras slutbetyget i provämnena dels på resultat på de nationella proven, dels på lärarnas bedömningar. Det senare innebär både att bedöma de delar som inte provas på proven och att bedöma vilken roll de nationella proven ska ha för betygssättningen. I svenska och engelska, där provbetyg inte finns idag, innebär det dessutom att avgöra vilken roll de olika delproven ska ha.

Enligt utredningsgruppens bedömning är detta ett system som gör betyg svåra att jämföra eftersom någon gemensam princip för hur de olika delarna ska sammanfattas i ett betyg inte finns. Delvis är det ofrånkomligt att det är på det sättet i det rådande systemet där ett lokalt fri utrymme ska finnas för att på bästa sätt främja elevens lärande. Samtidigt har dock betygen stor betydelse för elevernas framtida möjligheter. Ambitionen måste vara att i möjligaste mån skapa underlag för jämförelse utan att för den skull skapa total likformighet och därmed inskränka möjligheterna till att skapa goda lokalt förankrade lärandemiljöer.

Utredningsgruppen anser att ett steg mot ökade möjligheter till jämförbarhet är att det ska finnas ett provbetyg för de nationella proven som bestäms på samma sätt av alla lärare, inte som för närvarande att varje lärare själv bestämmer hur olika delprovsbetyg ska hanteras. Det senare innebär en form av godtycke och eftersom betygssättning är en myndighetsutövning bör utrymmet för godtycke vara så litet som möjligt.

Man kan naturligtvis diskutera hur de olika provdelarna ska vägas samman, men det som är viktigt för jämförbarheten är att sammanvägningen sker på samma sätt för alla och att alla vet hur den görs. Det kan jämföras med sättet att beräkna meritvärden. Viktningen av olika betyg (0 för IG, 10 för G osv.) kan ifrågasättas och diskuteras, men den har det obestridliga värdet att alla vet hur måttet är konstruerat. Detsamma gäller för ett provbetyg som konstrueras på samma sätt för alla elever enligt fastställda principer. Alla vet vad som gäller. Därigenom kan nationella provresultat jämföras och de kan också diskuteras i relation till slutbetygen på ett tydligare sätt än om tolkningen av delprovresultat är varje lärares ensak.

Det är dock viktigt att inte tolka ovanstående som att relationen mellan betyg på nationella prov och slutbetyg ska vara centralt fastlagd. Målet för skolans verksamhet, att skapa så goda lärandemiljöer som möjligt, får inte glömmas. Att ge proven en roll som examensprov skulle likrikta undervisningen på ett icke önskvärt sätt och står dessutom i uppenbar strid mot de nationella styrdokument. De slutliga betygen ska alltså fortfarande i enlighet med styrdokumentens skrivningar sättas av läraren, men man kan kräva av läraren att han eller hon redovisar hur resultaten på de nationella proven tolkats. Genom

Dnr 01-2003:2038

att betygen på dessa erhålls på samma sätt för alla elever kan de utgöra en gemensam referenspunkt för sådana tolkningar.

Sammanfattning och förslag

Som torde ha framgått finner utredningsgruppen det svårt att ge några enkla och entydiga förslag till svar på det uppdrag som Skolverket fått. I synnerhet uppdraget att bedöma provsystemets kvalitet har visat sig vara en mycket komplicerad fråga. Sammanfattningen av utredningsgruppens slutliga bedömningar samt gruppens förslag till Skolverkets ställningstaganden har lagts som inledning på rapporten under rubriken *Sammanfattning och förslag till Skolverkets bedömning*.

Appendix

Kunskapsbedömning och prov – en översikt

Den här bilagan har en teoretisk prägel syftar till att ge en översikt av olika aspekter av storskalig kunskapsbedömningen.⁴⁷

Inledning

Även om den teoriram och de resonemang som presenteras har en generell inriktning så ligger fokus på bedömning av teoretiska och ”akademiska” kunskaper. Det betyder att bedömningar av praktiskt kunnande, vilket naturligtvis är en stor uppgift inom många av gymnasieskolans program, intar en undanskymd roll i den här framställningen. Detta ska inte ses som ett uttryck för att bedömning av sådant kunnande skulle vara mindre viktigt. Det beror snarare på att den frågan inte behandlats i samma utsträckning som bedömning av teoretiska kunskaper av forskningen om kunskapsbedömning.

Det här avsnittet bör ses som en översiktlig skiss där flera problemområden endast berörs på ett ytligt och översiktligt sätt. Syftet är främst att ge ett underlag för och initiera en diskussion om mål, inriktning och utformning av kunskapsbedömning på nationell nivå.

Utländska erfarenheter kommer att ges betydande utrymme i ett försök att relatera vårt svenska system till mer internationella strömningar och tendenser. Dessutom är det naturligtvis så att den mesta forskningen inom området bedrivs utomlands, främst i USA, och därför kommer referenser att göras till utländsk forskning i rätt stor utsträckning.

Ett första väsentligt område gäller den begreppsapparat vi rör oss med och som inte alltid är så lättförståelig. Ett antal centrala begrepp och deras inbördes relationer kommer därför att lyftas fram och diskuteras.

Ett andra område gäller läroplaner, kursplaner, betygskriterier och standards och deras inbördes förhållanden.

Ett tredje område rör de instrument vi arbetar med, dvs. i första hand olika former av prov. En översiktlig beskrivning av deras teoretiska underlag, tekniska egenskaper, syften och användningsområden redovisas och diskuteras.

⁴⁷ Eftersom den huvudsakliga litteraturen på området är engelskspråkig och etablerade svenska begrepp ofta saknas kommer texten tyvärr att innehålla ett antal svåruthärdliga anglicismer.

Dnr 01-2003:2038

Några centrala begrepp

Området kunskapsbedömning innefattar många begrepp och termer, liksom många teorier, antaganden, metoder och instrument. Det är viktigt för den fortsatta diskussionen att försöka skapa en gemensam begreppslik bas och också att lyfta fram några centrala utgångspunkter för kunskapsbedömningens villkor och möjligheter.

Kunskapsbedömning, kunskapsutvärdering, pedagogiska mätningar, "assessment", prov, osv. är exempel på begrepp och termer som är ofta förekommer och som ofta skapar förvirring i diskussioner genom att de tilldelas olika innebörder. Ambitionen är att försöka lägga en gemensam referensram för våra vidare samtal.

Begreppet kunskapsbedömning är sammansatt av de två orden eller begreppen kunskap och bedömning. Begreppet kunskap har diskuterats av filosofer genom alla tider och det är givetvis inte möjligt att på något mer sofistikerat sätt fördjupa sig i frågan. I mer vardagliga sammanhang leder ordet kunskap lätt tanken mot något man har eller äger, en uppsättning fakta som kan uppfattas som en mängd eller ett pensum av kunskap som någon besitter. Detta är den kunskapsform som förekommer i olika typer av frågeprogram. "Vem skrev operetten Annie get your gun?" "När slutade andra världskriget?" Vilken EU-organisation ligger i Strasbourg?" osv.

I Lpo 94 är fakta en av fyra kunskapsformer som nämns och den form som enligt vissa av Skolverkets studier förefaller vara den i skolan (och bland allmänheten) dominerande tolkningen av innebörden i ordet kunskap. Denna form av kunskap kan ses som en passiv kunskapsform. Någon ställer en fråga och den svarande har eller har inte det rätta svaret.

Mot detta kunde man ställa ett mer aktivt kunskapsbegrepp, t.ex. kunnande. Detta implicerar att det handlar om kunskap i handling. Det är här mer fråga om den kunskapsform som kallas färdighet i läroplanen, men en form som ofta också förutsätter förståelse och förtrogenhet, de övriga två kunskapsformer som finns omnämnda i Lpo 94. Ett sätt att bryta pensumtänkandet kunde vara att ersätta ordet kunskap med kunnande. Man kunde då tala om att man bedömer kunnande i stället för att man bedömer kunskap. Ordet kunnande har dessutom en innebörd som ligger närmare det i skolsammanhang allt vanligare begreppet kompetens, även om detta begrepp inte heller har någon självklar innebörd.

Nationalencyklopedin anger några innebörder varav följande torde vara den i det här sammanhanget mest relevanta

kompetens (ytterst av senlat. *compete* 'ntia 'sammanträffande', 'överensstämmelse', av *co* 'mpeto 'sammanträffa', 'vara ägnad', 'vara kompetent', 'räcka till'), kunnighet, skicklighet

Begreppet kompetens har enligt denna beskrivning en påtaglig innebörd av kunskap i handling eller i varje fall en innebörd av beredskap för kunnigt handlande. Kompetens kan därmed ses som ett uttryck för en dynamisk kunskap medan kunskap i betydelsen pensum av fakta mer kan ses som en statisk kunskapsform. I begreppet kompetens skulle man också kunna lägga in förmågan att använda olika artefakter, dvs. manifestationer av mänskligt kunnande i form

Dnr 01-2003:2038

av olika verktyg, instrument och andra typer av fysiska eller mentala hjälpmedel. Kompetens skulle då inte bara vara kunskap (om) eller kunnande (i) som en individuell egenskap utan också en form av metakunnande i hur man använder ett kollektivt, manifesterat kunnande.⁴⁸

Ordet bedömning betyder enligt Svenska akademins ordbok

bedömning, bilda sig l. hysa en mening om (ngn l. ngt); värdera l. uppskatta l. betrakta (ngn l. ngt på ett visst sätt); avgiva sitt betänkande l. utlåtande om besaksheten af l. förtjänsten l. värdet hos, uttala sin mening om, fälla omdöme om, döma om. *Han har blifvit strängt, (o)gynnsamt, (o)mildt, (o)partiskt, välvilligt bedömd.*

Begreppet ger alltså klart uttryck för att någon betraktar och värderar någonting. Därmed kan man också säga att begreppet bedöma är närbesläktat med uttrycket utvärdera även om också detta begrepp ges många innebörder. Nationalencyklopedin ger följande definition

utvärderande, utvärdering, (noggrant) bedöma värdet av (resultatet av) ny verksamhet e.d. {**evaluera**}: ~ *projektet*; ~ *skattereformen*

Däremot är bedömning inte särskilt närbesläktat med begreppet mätning. Nationalencyklopedin definierar mätning på följande sätt:

mätning, serie åtgärder varmed en egenskap hos en företeelse, ett objekt eller ett ämne kan uttryckas med siffror på ett sådant sätt att relationer mellan empiriska företeelser representeras med relationer mellan tal. Mätning gör det möjligt att tillämpa matematik i empiriska sammanhang, t.ex. att formulera samband mellan variabler (storheter) som matematiska ekvationer, och att tillämpa kvantitativa metoder.

Pedagogisk mätning skulle således innebära att med hjälp av olika åtgärder, t.ex. prov, tilldela egenskaper (kunskaper) numeriska värden som kan hanteras matematiskt och statistiskt.

Bedömning och mätning behöver dock inte utesluta varandra. Mätningen innebär ”en serie åtgärder... varmed en företeelse kan uttryckas med siffror...”. Den genererar således data som kan behandlas med statistiska och matematiska metoder. Men när detta väl är gjort kan naturligtvis utfallet utsättas för bedömning. Mätningen kan således sägas vara en metod för att införskaffa information som sedan bedöms eller värderas i relation till någonting. Samtidigt är det givetvis så att den ”serie åtgärder” som vidtas vid själva mätningen måste baseras på någon form av (kvalitativ) bedömning. Man kan således säga att en mätning kräver bedömning för att vara meningsfull men en bedömning kräver inte mätning för att vara det.

Det mest svårfångade begreppet är det i engelskt språkbruk mycket vanliga uttrycket ”assessment”. Det har ingen klar svensk motsvarighet. Frågan är om det har det i det engelska språket. Collins English Dictionary ger följande beskrivning:

assessment, “1. the act of assessing... a student’s achievement on a course [...] 4. evaluation; estimation

⁴⁸ Det finns förvisso mycket mer att säga om begreppet kompetens. Det pågår också ett antal internationella projekt, t.ex. DeSeCo, definition and selection of competencies” (SFSO, 2003 eller Rychen & Salganik, 2003) med syfte att definiera vilka ”key competencies” som krävs i det nutida samhället och som i än högre grad förväntas krävas i det framtida.

Dnr 01-2003:2038

Där "assessing" baseras på "assess" som ges förklaringen

to judge the worth, importance, etc. of; evaluate

Med Collins definition skulle den mest närliggande översättningen vara att assessment är detsamma som kunskapsutvärdering på individnivå. Det skulle alltså vara själva bedömningsakten som är assessment. En lärare i ett klassrum som ständigt bedömer sina elevers kunskande skulle således ägna sig åt assessment.

Med utgångspunkt i denna definition skulle kunskapsbedömning på gruppnivå kunna kallas kunskapsutvärdering. PISA skulle i så fall kunna betecknas som utvärdering (evaluation) eftersom instrumenten inte ger underlag för bedömning av enskilda elever. PISA samlar också in bakgrundsinformation av olika slag vilket ger underlag för att kunna förklara och förstå utfall för olika undergrupper av elever, vilket stöder att studien karakteriseras som ett program för kunskapsutvärdering. Problemet är bara att PISA betyder Program for International Student *Assessment*.

Att tolka assessment som invidutvärdering stämmer inte heller med hur begreppet används i andra sammanhang. NAEP i USA betyder National *Assessment* of Educational Progress och är ett stort nationellt program för att mäta kunskapsutvecklingen inom olika ämnen. I det sammanhanget handlar det således inte heller om invidutvärdering utan om gruppjämförelser på statlig, regional eller annan nivå samt om kunskapsjämförelser över tid. Individjämförelser är överhuvudtaget inte möjliga med den valda tekniken för datainsamling. Såväl PISA som TIMSS använder för övrigt samma mättekniska ansats med så kallade "plausibla värden"⁴⁹ som NAEP och kan inte heller användas för individbedömning även om man så önskar. Alla tre program använder också enkäter för att samla in bakgrundsinformation som kan bidra till att förklara utfallet på kunskapsproven.

Den svenska nationella utvärderingen NU-03 liknar de internationella programmen genom sin kombination av kunskapsprov och insamling av bakgrundsdata. Även om NU-03 inte använder någon mer sofistikerad teknik för kunskapsmätning så innefattar projektet både kunskapsmätning och insamling av bakgrundsdata och NU-03 skulle då i konsekvens med PISA kunna betecknas som assessment, där assessment likställs med kunskapsutvärdering.

Det finns ytterligare exempel på förvirrande användning av assessment. En av de amerikanska motsvarigheterna till högskoleprovet har akronymen SAT, vilket fram till början av 1990-talet betydde Schoolastic Aptitude Test (vilket närmast kan översättas med "studieförmågetest"). I dag står A:et i stället för assessment och SAT betyder alltså Schoolastic Assessment Test (hur det nu ska förstås). I det här fallet handlar det om ett rent mätinstrument (test) som döpts om och getts ett namn mer i tidens anda. Mätning har blivit assessment.

CRESST (National Center for Research on Evaluation, Standards and Student Testing) i USA definierar helt enkelt assessment som "The process of gathering, describing, or quantifying information about performance". Med denna

⁴⁹ Statistiskt konstruerade mått på personers förmågor eller egenskaper.

Dnr 01-2003:2038

formulering undviker man att ta ställning till om det handlar om information på individ- eller gruppnivå och begreppet rymmer därmed alla former av informationsinsamling om "performance". Där "performance" får tolkas som det som eleven presterar eller visar.

Begreppet assessment skulle med denna definition kunna tolkas som en motsvarighet till mätning, men utan restriktionen att generera kvantitativa data. Assessment skulle därmed kunna ses som ett till mätning överordnat begrepp som kan innefatta mätdata (mätning), men därutöver även kvalitativ information som observationsdata, intervjudata, portfolios osv.

Assessment enligt ovanstående definition (CRESST) innefattar dock inte något värdeomdöme om den insamlade informationen. När ett sådant tillkommer kan de data som insamlats sägas utgöra underlag för utvärdering, eller - om det gäller värdering av data om kunskaper - om kunskapsbedömning.

Med CRESSTs definition av assessment som insamling, beskrivning eller kvantifiering av information som utgångspunkt borde även *uppföljning* vara ett svenskt begrepp som ryms inom assessment. Den insamling av betygsstatistik som Skolverket gör årligen skulle då kunna sägas vara en form av assessment. Denna information kan sedan få ligga till grund för värdering varvid man kunde tala om *utvärdering*.

Det man efter denna något omständliga genomgång kan konstatera är att assessment är ett mångtydigt och svårångat begrepp även i engelskt språkbruk. De kan således översättas med utvärdering (PISA), mätning (SAT), uppföljning (CRESST), eller individbedömning (Collins). Den mångtydigheten torde vi få leva med. Innebörden i begreppet får helt enkelt bedömas utifrån det sammanhang där det används.

Två andra begrepp som ofta förekommer i sammanhanget är *formativ* respektive *summativ* utvärdering (eller assessment i engelskspråkig litteratur). Formativ avser en fortlöpande utvärdering med syfte att ge underlag för att påverka den pågående processen, medan summativ utvärdering syftar till att ge ett sammanfattande omdöme när en verksamhet avslutats. Ett diagnostiskt prov är ett exempel på formativ utvärdering medan ett examsprov är ett exempel på summativ utvärdering.

I svenskt språkbruk kan kunskapsbedömning och kunskapsutvärdering i enlighet med ovanstående ses som två närliggande och till stor del överlappande begrepp. Om man vill definiera en skiljelinje bör den kanske främst knytas till objektet för bedömningen eller utvärderingen. Kunskapsbedömning i mer renodlad mening kan i så fall sägas ha fokus på bedömning av individers kunskaper eller kunnande. En lärares dagliga umgänge med elever innehåller således ständiga inslag av kunskapsbedömning.

Däremot kunde man tänka sig att begreppet kunskapsutvärdering får leda tanken till kunskapsbedömning av mer systematiskt slag, med inriktning mot grupper (klass, skola, kommun, hela riket eller en grupp nationer) och baserad på olika typer av instrument, främst olika typer av prov. Det primära syftet skulle då vara att från central nivå samla in och värdera information om elevers prestationer. Informationen går primärt från periferi till centrum.

Dnr 01-2003:2038

Samtidigt är ett viktigt syfte med de nationella proven att stödja lärarna i deras bedömning av kvaliteten i den enskilda elevens måluppfyllelse.⁵⁰ Informationen går då från centrum till periferi.

Sammanfattningsvis skulle resonemanget föra fram till att begreppet kunskapsbedömning i det här sammanhanget kunde förstås så att förledet kunskap i huvudsak utläses som kunnande eller kompetens. Det vill säga mer ses som kunskap i handling (eller en beredskap till sådan) än som kunskap i den mer allmänna betydelsen av pensum av fakta.⁵¹

Den andra delen – bedömning – innebär främst att ta ställning till och värdera det kunnande som olika instrument för mätning eller uppföljning ger underlag för att bedöma. I den metodiska och tekniska arsenalen bör således ingå verktyg för att inhämta både numeriska data (mätningar) och annan form av information ("assessment") liksom verktyg för att analysera denna och annan information (t.ex. från uppföljning), dra slutsatser och värdera (utvärdera) utfallet. Och dessutom ingår givetvis att tillhandahålla instrument (nationella prov och diagnostiska material) för att stödja lärarnas formativa och summativa bedömning av sina elever.

Styrdokument och standards

Att försöka sätta in de svenska styrdokumenten i ett internationellt perspektiv är inte helt enkelt. En översiktlig granskning av utvecklingen i ett antal (främst engelskspråkiga) länder visar dock på tydliga paralleller. Även om utgångsläget varit olika tycks det som om många trender och utvecklingslinjer går i samma riktning i de olika länderna. Det är väl å andra sidan knappast förvånande i en krympande värld där internationalisering och globalisering blir allt mer påtaglig.

Alla länder har någon form av övergripande mål för sin skolverksamhet, men i övrigt kan stora skillnader förekomma när det gäller hur centraliserad eller decentraliserad den mer konkreta målbeskrivningen är.

Vissa länder har av tradition haft starkt decentraliserade skolsystem t.ex. USA och Storbritannien. Där har trenden under 1990-talet varit en ökande central styrning av skolverksamheten genom en ökande betoning av nationella mål och vikten av måluppfyllelse. Andra länder, t.ex. Sverige, har gått i motsatt riktning och minskat den centrala regelstyrningen till förmån för ett ökat lokalt inflytande. I båda fallen kan dock målstyrning sägas ha varit den övergripande principen. Skillnaderna ligger snarast i vilken betoning som lagts på resultatuppföljning och ansvarsfördelning.⁵²

⁵⁰ Eftersom de nationella proven inte är examensprov avses detta stöd ske på ett mer indirekt sätt.

⁵¹ Därmed dock inte sagt att kunskapsformen fakta inte skulle vara väsentlig. Det handlar mer om var tyngdpunkten läggs än om att utesluta den ena eller andra typen av kunskap.

⁵² En annan skillnad ligger i att regelstyrning anger vad undervisningen skall innehålla (och eventuellt hur den ska genomföras) medan målstyrning anger vad eleven skall kunna efter genomgången utbildning. Detta är som bekant inte samma sak.

Dnr 01-2003:2038

Standards

En annan påtaglig (och kanske följdénlig) trend har varit ”The standards movement”. Den har inneburit ett markant ökat intresse från nationell nivå att fastlägga ”standards” för utbildningssystemen. Att enbart veta om den ena gruppen varit bättre eller sämre än den andra har inte ansetts tillräckligt utan ambitionen har varit att fastställa *vad* eleverna ska kunna för att de ska kunna anses kunna tillräckligt, dvs. man har velat formulera standards för vad som är godtagbara kunskaper på olika nivåer (år 5 och år 9 för svensk del).⁵³

Chatterji (2002) beskriver i en litteraturöversikt den amerikanska utvecklingen på följande sätt (sid 346.)

The concept of systemic reform in U.S. education emerged in the 1980s [...]. The defining components of systemic reform were threefold

1. The establishment of challenging standards in the academic disciplines that would define what students should know and be able to do;
2. Alignment⁵⁴ of curriculum and instruction, assessment and accountability, and teacher certification and professional development with the new academic standards; and
3. Revamping of school governance structures, allowing schools and teachers greater autonomy in how they organized instructional programmes to achieve high standards of student performance at the local level.

Likheterna med det nuvarande svenska systemet med deltagande målstyrning, ökat inflytande för professionen och ett lokalt friutrymme är slående.

De två typer av standards som det oftast talas om är ”content standard” och ”performance standard”. Phillips (1994) beskriver dem på följande sätt:

A content standard represents what the students are expected to learn. It often gets operationalized in terms of curriculum guides, becomes the subjects of textbooks, and represent the goals and objectives of teaching and assessment. However, a distinction is useful between the range of things students are expected to learn (*content standard*) versus the procedures used to demonstrate the attainment of acceptable levels of learning (*performance standards*). (s. 194)

En jämförelse med svenska kursplaner visar att den närmaste motsvarigheten till content standards är ”mål att uppnå” medan performance standards motsvaras av betygskriterierna. Detta framträder tydligt i National Educational Goals Panel (1993) s. 22, där man kan läsa

Performance standards are *not* the skills and modes of reasoning referred to in the content standards. Rather, they indicate *both* the nature of the evidence (such as an essay, mathematical proof, scientific experiment, project, or combination of these) required to demonstrate that content standards have been met *and* the quality of the student performance that will be deemed acceptable (what merits a passing or an A grade) (kursiverat här).

⁵³ Det finns en omfattande litteratur om standards, se t.ex. den litteraturöversikt som utgetts av utbildningsministeriet i British Columbia (Ministry of Education British Columbia, 2002)

⁵⁴ ”Alignment” är ett viktigt begrepp, som ofta betonas i litteraturen om standards, och kan väl närmast översättas med ”att vara i linje med”.

Dnr 01-2003:2038

Detta stämmer väl med den svenska synen att kursmålen anger vad eleven ska kunna när kursen är avslutad medan betygskriterierna talar om på vilka sätt eleven kan visa de kvaliteter i måluppfyllelse som berättigar till olika betyg.⁵⁵

Om man tittar på content standards i t.ex. USA, Kanada eller Australien, eller ”program of study”, som är den engelska motsvarigheten, finner man vid en jämförelse med de svenska mål att uppnå att innehållet i de utländska motsvarigheterna i allmänhet är betydligt mer specificerat. I svenska kursplaner anges innehållet i de flesta ämnen i mycket generella ordalag. De liknar närmast det som i engelskspråkig litteratur ofta benämns ”content strands”, vilket närmast kan översättas med kunskapsområden.

Exemplet matematik

I Sverige har vi en uppdelning av målen i två skilda kategorier, mål att sträva mot och mål att uppnå. Tanken är att mål att sträva mot ska ligga till grund för planeringen av undervisningen medan mål att uppnå ska ligga till grund för bedömningen och betygssättningen av elevernas prestationer.

I matematik kan man i mål att sträva mot t.ex. läsa (Grundskolan 2000)

Strävan skall alltså vara att eleven utvecklar sin tal- och rumsuppfattning samt förmåga att använda

- grundläggande talbegrepp och räkning med reella tal, närmevärden, proportionalitet och procent,
- olika metoder, måttssystem och mätinstrument för att jämföra, uppskatta och bestämma storleken av viktiga storheter,
- grundläggande geometriska begrepp, egenskaper, relationer och satser,
- ---

I ”mål som eleven ska ha uppnått i slutet av nionde skolåret” formuleras målen på t.ex. följande sätt

Eleven skall ha förvärvat sådana kunskaper i matematik som behövs för att kunna beskriva och hantera situationer samt lösa problem som vanligen förekommer i hem och samhälle och som behövs för fortsatt utbildning.

Inom denna ram skall eleven

- ha utvecklat sin taluppfattning till att omfatta hela tal och rationella tal i bråk- och decimalform,
- ha goda färdigheter i att kunna använda överslagsberäkning och räkning med naturliga tal och tal i decimalform samt procent och proportionalitet i huvudet, med hjälp av skriftliga räknemetoder och med tekniska hjälpmedel,
- ---

Det är svårt att här se någon principiell skillnad i graden av konkretion. Mål att sträva mot talar i futurum medan mål att uppnå uttrycks i imperfektum. I övrigt är skillnaderna små. Betygskriterierna å sin sida är uttryckta i presens.

Matematik är för övrigt inte det ämne som är otydligast när det gäller det konkreta innehållet i ämnet eller kursen, snarare tvärtom.

⁵⁵ Det gäller dock inte för betyget Godkänd i år 9 där mål att uppnå (content standards) sammanfaller med kriterierna för samma betyg (performace standards), vilket knappast bidrar till att tydliggöra skillnaden mellan vad som är mål och vad som är kriterier, dvs. beskrivningar av vad en elev ska visa för att manifestera måluppfyllelse av viss kvalitet.

Dnr 01-2003:2038

Tanken i systemet är att de i generella termer formulerade kursmålen ska konkretiseras i det lokala kursplanarbetet. Det är således lärarna tillsammans med eleverna som ska bestämma vad det innebär att ha "en utvecklad taluppfattning" och att ha "goda färdigheter i att kunna använda överslagsberäkning". Några exempel på vad måluppfyllelsen innebär, vilket är vanligt i utländska måldokument eller content standards – eller i kommentarer till dessa – finns inte i de svenska styrdokumenterna.

Olika nivåer

Performance standards är i USA ofta uttryckta i en fyrgradig skala och betecknas med termerna "Below basic", "Basic", "Proficient", eller "Advanced", vilket kan jämföras med vår svenska betygsskala. "Below basic" används ibland inte. Detta motsvarar då närmast att inte få betyg i ämnet eller att ej uppnå målen (EUM), vilket är det uttryck vi använder i Sverige för elever i grundskolan som inte är godkända.

I England har man flera steg i sina performance standards. "Attainment targets"⁵⁶, som de kallas, har åtta olika nivåer från level 1 till level 8 samt en nionde nivå för proficient.

Standards av båda slagen är ofta fastlagda för olika stadier eller "key stages" som innefattar ett större eller mindre antal årskurser, i England i allmänhet tre, i USA mer varierande. I Sverige kan vi sägas ha två nivåer eller stadier med formulerade mål, en upp till och med år 5 och en från år 6 till 9.

I Sverige kan man alltså säga att våra mål att uppnå och betygskriterier närmast motsvarar de två typer av standards som är vanliga i de anglosaxiska länderna, även om mål att uppnå i många fall mer liknar content strands än content standards. I de anglosaxiska länderna är dock standards snarare kopplade till nationella eller (del)statliga utvärderingar än till betyg och betygssättning.

Instrument som baseras på standards

Flertalet av de moderna internationella studierna av elevers kunskaper baseras på standards.

National Assessment of Educational Progress (NAEP)

Stor betydelse för den amerikanska utvecklingen av standards har det tidigare nämnda projektet NAEP (National Assessment of Educational Progress) haft. NAEP är ett nationellt utvärderingsprogram i USA med syfte att jämföra kunskapsutveckling mellan regioner, stater mm. och att ge en bild av kunskapsutveckling över tid. Programmet startade i början av 1970-talet och var då ett typiskt normrelaterat prov (mer om vad detta innebär senare). Det har i olika avseenden haft stor betydelse för såväl den tekniska som den begreppsliga utvecklingen av storskaliga mätinstrument.

Från början var proven strikt normrelaterade med en fix normgrupp. Detta ledde så småningom till problem genom att det kunnande eleverna visade på pro-

⁵⁶ Skulle väl närmast översättas som mål att uppnå.

Dnr 01-2003:2038

ven ökade och i slutet av 1980-talet visade det sig att samtliga stater låg över medelvärde (den s.k. Lake Wobegon effekten).⁵⁷ Detta ledde till en omfattande diskussion om meningen med normrelaterade prov och så småningom bestämdes att projektet i stället skulle baseras på standards. Sådana infördes 1992.

Dessa standards för olika ämnen har sedan fått stor betydelse för utvecklingen av standards i olika delstater. NAEPs övergripande standards är fastställda i samverkan mellan olika professionella och officiella organisationer men ska alltså inte ses som läroplaner eller kursplaner. Däremot har de utgjort utgångspunkter för konstruktionen av läroplaner, kursplaner och nationella eller andra prov.

När projektet inleddes på 1970-talet var rädslan för federal inblandning i delstaternas angelägenheter så stor att några resultat på delstatsnivå inte redovisades. Efterhand har dock denna oro mildrats och i dag redovisas resultaten för varje delstat och det finns också möjligheter att få resultatredovisningar på ännu lägre nivå.

Skälen till att standards införts i ett sådant nationellt utvärderingsprogram var att det inte ansågs tillräckligt att endast jämföra hur olika regioner och stater presterade i relation till varandra, utan man ville också veta hur elever presterade i relation till fastställda kunskapsinnehåll och kunskapskvaliteter, dvs. i relation till content standards och performance standards. Det är givetvis väsentligt varhelst sådana standards används att de är förankrade hos olika intressegrupper och att de får hög legitimitet, men de utgör alltså i många fall inte någon fastställd läroplan eller kursplan även om det i allmänhet råder god överensstämmelse med de olika staternas läroplaner och kursplaner och de standards som ingår i överstatliga utvärderingsinstrument.⁵⁸

PISA

PISA kan, som tidigare antytts, ses som en internationell motsvarighet till NAEP. PISA marknadsförs som en icke läroplansbunden studie. I stället hävdas att studien mäter kunskaper och kompetenser som krävs i det nutida samhället.⁵⁹

Även PISA kan sägas ha anammat tänkandet i standards. När det gäller läsförståelse finns det tre content strands som definierats som "information", "tolkning" och "reflektion". Dessa preciseras sedan i ett antal exempel på "continuous" och "non-continuous" texter, vilka tillsammans med de tre kunskapsområdena kan betecknas som content standards. Det finns ytterligare specifikationer

⁵⁷ Detta visade sig så småningom vara en effekt av att samma uppgifter användes, vilket fått till följd att lärarna undervisade men inriktning mot provet, "teching to the test", som det brukar kallas.

⁵⁸ Det finns dock exempel på stater där det råder betydande skillnad mellan NAEPs standards och statens egna standards. I Louisiana bedömdes 7 procent av fjärdeklassarna som proficient enligt NAEP mot 80 procent enligt statens egen utvärdering. (*NASBE: A primer on state accountability and large-scale assessment.*)

⁵⁹ Eftersom man mäter kunskaper i de uppenbara skolämnena läsförståelse, matematik och naturvetenskap torde det dock vara ofrånkomligt att kursplaneutformningen i olika länder har stor betydelse för utfallet.

Dnr 01-2003:2038

samt exempel på uppgifter för att tydliggöra vad som avses (se PISA 2000, s 18 ff).

Som performance standards har fem olika nivåer definierats från den lägsta "Level 1" till den högsta "Level 5". Dessa är formulerade i verbala utsagor om vad en elev inom respektive nivå förväntas kunna (se t.ex. PISA 2000, Skolverkets rapport 209, s.22). Några exempel på metoder för att bestämma sådana nivåer beskrivs senare.

En jämförelse av de svenska PISA-deltagarnas betyg med resultaten på läsförståelseskalan visar att G-elever fördelar sig kring gränsen mellan nivå 2 och 3, VG-eleverna kring gränsen mellan 3 och 4 och MVG-eleverna kring gränsen mellan 4 och 5. Cirka hälften av MVG-eleverna ligger i nivå 5 och hälften i nivå 4. Intervallen mellan olika betyg och mellan PISAs nivåer överensstämmer förvånansvärt väl trots att betygen sätts enligt helt andra principer än de som gäller för PISAs nivåindelning. Givetvis finns det också betydande överlappning mellan olika nivåer men samstämmigheten är slående (Skolverket 2001, figur 3.11).

Standards, betyg och likvärdig bedömning

Betygssättningen kan vara kopplad till standards (som i Sverige om vi betraktar mål att uppnå och betygskriterier som standards) men behöver inte vara det. I USA är betygssättningen ofta en lokal angelägenhet, och i hög grad provbaserad. Ett antal privata företag tillhandahåller examensprov och i vissa stater finns det också centralt utformade prov. Betygssättningen behöver som nämnts inte vara direkt baserad på nationella standards även om sådana kan ligga till grund både för utformningen av statliga och lokala kursplaner. Standards kan alltså ligga till grund för utvärderingen av skolan även om de inte ligger till grund för betygssättningen.

I många länder finns dock liksom i Sverige nationella prov som baseras på de standards som definieras i nationella kursplaner. Så är t.ex. fallet i Storbritannien där det finns nationella "curriculum assessment". Även i Kanada och Australien förekommer liknande prov. Deras koppling till standards och betygssättning är dock inte helt lätt att utvärdera för en extern betraktare.⁶⁰

Ett begrepp som ägnats mycket uppmärksamhet vid diskussioner om standardsbaserade system, och som är kopplat till frågan om likvärdighet, är "alignment". Det är en fråga som vi i Sverige kan känna igen. Chatterji (2002) ställer den på följande sätt:

Does the meaning of content standards remain the same across levels of the state-district-classroom organizational hierarchy when local reforms are implemented. (sid. 363).

Denna problematik kan vara grunden till att så många forskare betonar betydelsen av klara content standards.

⁶⁰ Viss information, främst med inriktning mot "reading literacy", finns i *PIRLS 2001 Encyclopedia* (Mullis m.fl., 2002). Även INCA, International Review of Curriculum and Assessment Frameworks Archive tillhandahåller information om olika länders system på sin hemsida <http://www.inca.org.uk/>

Dnr 01-2003:2038

Sammanfattningsvis kan man säga att Sverige följt med den trend som sedan 1980-talet inneburit ett starkt ökat intresse för att formulera standards för utbildningen innehåll och kvalitet. Det kan också uttryckas som att det funnits en trend mot ökad målstyrning. Sverige har dock intagit en viss särställning genom att målbeskrivningarna i våra kursplaner är mycket generellt hållna och därför ger liten vägledning när det gäller det konkreta innehållet i undervisningen och vad en elev ska kunna inom olika kunskapsområden. Man kan säga att i jämförelse med många andra länder saknas den nivå som motsvarar content standards. De nationella kursplanerna anger snarast content strands och performance standards.⁶¹

I Sverige finns knappast några starka professionella organisationer som har resurser att på eget initiativ utforma mer konkreta standards på det sätt som skett i USA i t.ex. matematik genom NCTM (National Council of Teachers of Mathematics), utan lärarna förväntas själva komma fram till konkretioner av de kunskapsområden som skisseras i grova drag i de nationella kursplanerna.

Detta leder till frågan vad en svensk lärare ska utgå från när han eller hon väljer innehåll i undervisningen. Några närmare beskrivningar eller exempel ges inte i de nationella styrdokument. Det finns inte heller några nationella ämnesföreningar som ger rekommendationer eller riktlinjer för lämpligt innehåll. Det torde därmed, i varje fall i vissa ämnen, vara läroböckernas val av innehåll som är mest styrande för innehållet i undervisningen. I andra ämnen dominerar däremot i högre grad det innehåll man själv i samverkan med sina kollegor (och eleverna) kommer överens om (se t.ex. Skolverkets kvalitetsgranskningar).

Många ämnen har också en stark ämneskanon som gör att ämnesinnehållet ses som rätt givet och självklart, något som kanske främst gäller de naturvetenskapliga ämnena. Men vilka bevekelsegrunder lokala val vilar på är oklart och det är svårt att se hur likvärdighet ska kunna uppnås i ett system där ingen central aktör har eller tar ansvar för att peka ut någon viss inriktning när det gäller innehållet i ämnet.

Det nationella instrument (eller dokument) som i nuvarande svenska system framstår som normgivande för innehållet är de nationella proven. De definierar, konkretiserar och exemplifierar innehållet i ämnet (content standards) och de olika kunskapskvaliteter som olika betygskriterier representerar (performance standards). Därmed kan de ses som komplement till och exemplifieringar av styrdokumentens kursplaner och betygskriterier. De nationella proven kan således ses som en del av skolans styrsystem genom att de förtydligar och konkretiserar mål och betygskriterier. Detta är givetvis väsentligt för en likvärdig bedömning av elevernas prestationer. Hur en likvärdig betygssättning ska uppnås i ämnen som saknar normgivande nationella prov är mer oklart.

Öppna eller slutna styrdokument?

När det gäller den mer allmänna frågan om vilken typ av läroplan eller kursplan - en öppen och ospecificerad med stort lokalt inflytande (som den svenska)

⁶¹ Den internationella bilden är dock långt ifrån entydig och det finns exempel på content standards som liknar de svenska mål att uppnå när det gäller konkretionsgrad.

Dnr 01-2003:2038

eller en mer detaljerad och centralt föreskrivande (som i England) – som är bättre för elevernas lärande är bilden oklar. Någon särskilt omfattande forskning finns inte inom området.

I en Nya Zeeländsk kunskapsöversikt (Carr m.fl. 2001) refereras en studie av Black och Atkin (1997) som baseras på TIMSS-data från 45 länder och OECD-data från 13 länder. Någon klar bild var svår att se. Av de tio länder som hade de bästa resultaten hade åtta centralt utformade kursplaner medan två baserade undervisningen på lokalt utformade kursplaner. Men förhållandet var detsamma bland de tio länder som lyckades sämst. Även där hade åtta nationellt föreskrivande kursplaner. En av författarna som har god kännedom om England och Wales hävdar utifrån sina erfarenheter att det finns få bevis för att centraliseringen skulle förbättrat lärandet, men mycket som talar för att effekterna varit negativa, främst beroende på att det hade väckt motstånd hos lärarna.

Det finns således enligt denna studie inga belägg för att en tydlig och detaljerad nationell läroplan skulle höja kunskapsnivån generellt i skolan. Däremot torde man kunna hävda att tydliga riktlinjer har betydelse för rättvisan och likvärdigheten i bedömningen av den enskilda individen. Den som bedöms bör veta vad som bedöms, hur det bedöms och vad han eller hon förväntas visa.⁶²

Prov

Målen för skolans verksamhet formuleras i olika styrdokument (läroplaner och kursplaner) och formuleras på många sätt: som mål, som standards, som betygskriterier eller på annat sätt. För att ta reda på i vilken utsträckning de uppsatta målen uppnås måste emellertid information insamlas och bedömas på något systematiskt sätt.⁶³

Det tveklöst vanligaste sättet att samla information och att pröva måluppfyllelse inom utbildningssystem är individuella skriftliga prov. I varje fall är det så när det gäller äldre elever och mer storskaliga projekt (även om det också finns

⁶² Det finns naturligtvis mycket att säga om utformningen av standards och deras konsekvenser. I början av 1970-talet fanns i Sverige ett stort projekt kallat MUT (målbestämning och utvärdering). Det gick i tidens anda, och som ett arv av 1960-talets behavioristiska och utvecklingsteknologisk strömningar, ut på att reducera varje kunskapsmål till en minsta odelbar kunskapsatom som kunde uttryckas i beteendetermer. Detta resulterade i ständigt växande målkataloger som till slut sprängde alla ramar och projektet lades 1975 ner över en natt av dåvarande SÖ-chefen Jonas Orring. Idag tror ingen pedagogisk forskare på en sådan grad av målprecisering. Det är numera allmänt accepterat att det måste finnas ett professionellt friutrymme vid formuleringen av mål och standards. Någon total precision kan inte uppnås när det gäller att bedöma något så komplicerat och mångskiftande som det mänskliga kunnandet. Det handlar om mänsklig bedömning och om att ge ett så gott stöd som möjligt till denna bedömning. Uppgiften är således att hitta en grad av målbeskrivning som är optimalt tydlig utan att bli föreskrivande. En inte helt lätt uppgift.

⁶³ Den form av prov som diskuteras är de som på engelska kallas "achievement test", dvs. prov som avser inhämtade kunskaper eller kompetenser, inte prov av typen "aptitude test", som främst handlar om att testa "begåvning" eller (studie)potential. Högskoleprovet kan t.ex. ses som en form av aptitude test medan nationella prov kan betraktas som mer renodlade kunskapsprov. Varje form av prov ha dock inslag av både kunskaper och potential. Det handlar mer om var tyngdpunkten är förlagd.

Dnr 01-2003:2038

exempel där laborationsuppgifter, processtudier, videostudier och liknande ingår).

Ordet prov har sin grund i prövning, att pröva någons förmåga till någonting. I skolan gäller det vanligen elevers förmåga att ge svar på (fakta)frågor, att lösa problem av något slag eller att beskriva, diskutera och analysera något förhållande. Det vanligaste är att prövningen genomförs med papper och penna och under övervakning så att inga otillåtna hjälpmedel kan användas. Även i övrigt brukar prov genomföras under standardiserade villkor för att ge så rättvisa (likformiga) förutsättningar som möjligt.

Muntliga prov har dock funnits under lång tid, t.ex. i form av husförhör, muntliga studentförhör, examensprövningar vid universitet osv. Även i dag är de vanliga. Nackdelar kan vara att de är tidskrävande och att de kan uppfattas som subjektiva och orättvisa. De kan också svåra att analysera när det gäller giltighet (validitet) och tillförlitlighet (reliabilitet). Ytterligare en nackdel – i varje fall om det gäller examinerande prov – kan vara att den som provas är utlämnad till den som genomför prövningen. Eftersom dokumentation oftast saknas kan det vara svårt att i efterhand bedöma vad som sagts. Muntliga prov kan dock ha fördelar genom att de ger möjligheter till flexibilitet, improvisation och uppföljning. Idag betonas dessutom olika former av kommunikativ förmåga alltmer och inte minst för att framhålla detta bedöms det som angeläget att de svenska nationella proven innehåller en muntlig del.

Skriftliga prov har en lång tradition. Det har utvecklats olika teorier kring hur de bör konstrueras, vad de kan och inte kan mäta och vilka tekniska egenskaper de bör uppvisa för att vara av god kvalitet. Det vanligaste är att prov genererar numeriska värden (poäng), dvs. kan ses som mätningar, och att de baseras på item. Dessa item kan vara mer eller mindre omfattande och generera mer eller mindre många poäng. Ytterligheterna kan sägas vara å ena sidan flervalfrågor som ger en poäng per fråga, å andra sidan uppsatser som bedöms holistiskt med ett sammanfattande omdöme om hela arbetet. Det senare utesluter dock inte att en uppsats bedöms ur ett antal aspekter som redovisas var för sig, eventuellt med poängsättning inom varje aspekt. En sådan uppdelning syftar till att göra bedömningen mer likvärdig och reliabel genom att ett antal avgränsade aspekter beskrivs med hjälp av preciserade rättningsanvisningar ("scoring rubrics"). Mot detta kan ställas en mer holistisk bedömning där helheten ses som något annat än summan av delarna. I det senare fallet får man i stället utgå från jämförelser ett antal "typuppsatser" med utförliga kommentarer, vilka bedömts representativa för olika standards eller betyg.

Poängbedömda prov är goda exempel på mätningar som grundar sig på bedömningar. Hur många poäng en viss uppgift ska kunna ge är således en ren bedömningsfråga, även om den kan baseras på mer eller mindre välutvecklade teoretiska grunder, t.ex. i form av olika taxonomier (Bloom, SOLO etc.).

Under senare år har det alltmer talats om att proven ska vara verklighetsanknutna och att uppgifterna ska vara realistiska för att pröva elevers förmåga att hantera verkliga situationer. I engelskt språkbruk talas om "performance assessment" eller "authentic assessment". Detta är naturligtvis en sanning med betydande modifikation. En provsituation i skolan är alltid i viss mening artifi-

Dnr 01-2003:2038

ciell. Eleven sätts att lösa konstruerade problem, oftast beskrivna i text och bild, under specifika villkor som snarast är motsatta dem som gäller i en realistisk situation. Oftast får man inte använda de hjälpmedel man skulle använda om det handlade om problemlösning utanför skolan. Man får inte heller rådfråga andra människor på det sätt man gör i en autentisk situation. Vad man möjligen kan säga är att man försöker simulera en realistisk situation och begär att eleven ska agera under ett antal icke-realistiska restriktioner.

Validitet och reliabilitet

Varje provform har sina förutsättningar och kan analyseras med avseende på sina tekniska kvaliteter, t.ex. beträffande reliabilitet och validitet. Validiteten är oftast knuten till i vilken utsträckning proven, i relation till fastställda mål eller standards, prövar relevanta kunskaper och förmågor och om de ger eleven möjligheter att visa sina kunskaper på ett lämpligt sätt. Enligt moderna synsätt är dock validitet inte enbart en egenskap hos provet utan är också kopplad till hur provresultaten används och vilka konsekvenser de har för provdeltagaren. (Messick 1989).

Reliabiliteten uttrycker i vilken utsträckning olika slumpfaktorer påverkar resultatet. Ett fullständigt oreliabelt prov tilldelar eleverna poäng på samma sätt som om de singlat slant om svaren. Ett fullständigt reliabelt prov saknar därmed helt inslag av slump, men sådana prov finns inte. Allmänt kan man säga att ett prov blir mer reliabelt ju längre det är och ju fler uppgifter det innehåller. Samtidigt kan ett prov av praktiska och andra skäl inte vara hur långt som helst och eleverna kan inte heller sitta och göra prov hur länge som helst. Alla prov är därmed behäftade med mer eller mindre stora mätfel.

Ambitionen är dock att inom givna ramar få så hög reliabilitet som möjligt, dvs. så litet slumpinflytande som möjligt. Det finns olika tekniker för att konstruera kvantitativa mått på reliabilitet, framför allt för itembaserade prov (KR-20, Coefficient alfa etc.).

I de fall olika bedömare ska ta ställning till olika svar och bedöma hur många poäng (eller vilket betyg) de ska ge behöver också samstämmigheten mellan olika bedömare (den så kallade interbedömarreliabiliteten) prövas. För detta finns det metoder och tekniker.

Vid konstruktionen av prov kan olika vikt läggas vid olika egenskaper. Tidigare ungefär fram till 1990-talet prioriterades mätsäkerheten (reliabiliteten) och man brukade säga att utan reliabilitet ingen validitet. Under 1990-talet har dock validiteten kommit att prioriteras i högre utsträckning och delvis på bekostnad av reliabiliteten. I varje fall har det blivit svårare att konstruera mått på reliabiliteten. Det beror till stor del på den tidigare nämnda ambitionen att konstruera mer autentiska och realistiska uppgifter som ger utrymme för varierande svar, och där svaren inte ens alltid är i förväg givna utan kan bero på vilka antaganden som görs av den som löser uppgiften.

Ett annat sätt att uttrycka samma sak är att säga att proven till stor del har övergått från att vara kvantitativt baserade till att vara underlag för kvalitativ bedömning. I stället för att konstruera numeriska mått på validitet och reliabilitet

Dnr 01-2003:2038

får man utgå från bedömningar av relevans och giltighet (validitet), och tillförlitlighet (reliabilitet).

Vad som är väsentligt, enligt t.ex. de standards för provkonstruktion som fastställts av de amerikanska organisationer som sysslar med provverksamhet, ETS, APA och NCME,⁶⁴ är dock att tillvägagångssätt och utfall vid fastställande av validitet och reliabilitet redovisas. Inte minst anses detta vara viktigt av rättsäkerhetsskäl. Det betonas också att bedömningen inte får baseras på ett enskilt prov utan att det bör finnas flera underlag. Man talar i ökande utsträckning om betydelsen av "Multiple Measures".⁶⁵ Detta kan ses som ett sätt att öka tillförlitligheten i bedömningen, en form av bedömningstriangulering.

Klassisk och modern testteori

När det gäller testteori talas vanligen om två former, klassisk och modern. Den klassiska började utvecklas i början av 1900-talet och var mer eller mindre fullbordad på 1960-talet. Den bygger på antagandet att en provdeltagares provpoäng kan ses som summan av en "sann" poäng och ett mätfel. Utifrån detta grundantagande och några andra antaganden som t.ex. gäller slumpmässighet och korrelationer mellan mätfel och poäng kan formler för beräkning av reliabilitet och mätfel bestämmas.

Den klassiska testteorin ställer svaga villkor och är därmed giltig i de flesta provsammanhang. Den har dock nackdelar. Till exempel kan man inte avgöra om uppgifternas eller provens egenskaper beror på den aktuella provgruppen eller det aktuella provet. Uppgifternas egenskaper (t.ex. mått på uppgifternas svårighet och diskriminering, dvs. lösningsproportioner och biseriella korrelationer) beror både på provet och på gruppen, liksom provets egenskaper i form av reliabilitet och mätfel.

För att eliminera denna effekt försöker man skapa prov som är så likartade (parallella) som möjligt. Det innebär att man försöker ha lika många uppgifter i proven, försöker se till att de är av samma typ och svårighetsgrad och att de har så likartat innehåll som möjligt. Därigenom antas proven bli mer jämförbara och eventuella skillnader i lösningsproportioner eller poängfördelningar kan då hänföras till de prövade grupperna.

Den moderna testteorin, "item response theory (IRT)", bygger på starka antaganden om endimensionalitet och lokalt oberoende.⁶⁶ Detta är nödvändig villkor för att de statistiska modeller som teorin bygger på ska vara giltiga. Villkoren är aldrig helt uppfyllda men teorin anses ha så stora fördelar i övrigt att den kommit att få stor betydelse vid stora undersökningar av typen NAEP, TIMSS, PISA och PIRLS.

Fördelarna är framför allt att de deltagandes prestation (förmåga) och uppgiftens egenskaper (svårighetsgrad) kan anges på samma skala. Det är vidare möj-

⁶⁴ ETS (Educational Testing Service), APA (American Psychologist Association), NCME (National Council for Measurement in Education).

⁶⁵ Se t.ex. temanummer av *Educational Measurement: Issues and Practice*, 2003, Volume 22, Nr 2.)

⁶⁶ Det finns också mer komplexa flerdimensionella modeller.

Dnr 01-2003:2038

ligt att skatta provdeltagares resultat på uppgifter de inte gjort, vilket gör att man kan ge olika elevgrupper olika provhäften men skatta resultat för samtliga elever på samtliga uppgifter. Detta förutsätter dock antingen att några uppgifter är gemensamma för samtliga häften eller att häftena på annat sätt länkas till varandra.

En ytterligare fördel är att det inte som i klassisk testteori är nödvändigt att prov är parallella för att resultat ska bli jämförbara. Förutsatt att det finns vissa gemensamma uppgifter (ankaruppgifter) är enskilda provtagares resultat på olika prov jämförbara (kan placeras på samma skala) även om proven inte är parallella.

En nackdel med den moderna testteorin är att den är statistiskt och tekniskt komplicerad. Teorin utvecklades redan på 1940-talet men den kräver datorer för att kunna tillämpas och började därför inte utnyttjas praktiskt förrän på 1970-talet.

En annan nackdel är att resultaten presenteras i form av latent värden, dvs. ”poäng”⁶⁷ som genereras av den använda statistiska modellen. Rangordningen mellan elever kan då bli en annan när man baserar den på latent poäng än när man baserar den på observerade (manifesta) poäng på provet. Femton poäng på ett prov kan alltså ge en högre latent poäng än sjutton poäng på samma prov. Detta har att göra med att i IRT har det betydelse *vilka* uppgifter man besvarat rätt. I klassisk testteori ger tre rätt besvarade uppgifter (i ett prov med t.ex. tjugoen poängsuppgifter) tre poäng oberoende av vilka av uppgifterna som besvarats korrekt. I IRT däremot innebär vanligtvis varje kombination av tre rätt olika resultat. Rätt svar på uppgifterna 1, 2 och 3 kan ge ett annat resultat än rätt svar på uppgifterna 1, 2 och 4. Detta är svårt att förklara för t.ex. lärare, elever och föräldrar. Tillsammans med teorins tekniska komplexitet medför detta att IRT i skolsammanhang i stort sett endast används i stora studier där intresset gäller jämförelser på gruppnivå.

Normrelaterade prov

Den klassiska och moderna testteorin gäller alltså främst ett provs egenskaper som mätinstrument. Reliabiliteten kan sägas vara ett typiskt exempel på ett mått på ett provs tekniska eller instrumentella kvaliteter. Den klassiska testteorin ger mindre precisa mått i form av genomsnittsvärden på t.ex. reliabilitet och mätfel medan den moderna testteorin är mer avancerad och ger motsvarande mått som funktioner av (latent) poäng.

En annan fråga av stor betydelse är hur de resultat proven genererar ska användas. Ska de användas för urval till en utbildning? Är det viktigaste att få eleverna rangordnade på ett korrekt sätt? Ska provet användas för betygssättning? Ska de ge en så god bild som möjligt av kvaliteterna i elevernas kunnande?

I det förstnämnda fallet, där det främst handlar om urval och rangordning, talar man om grupprelaterade eller normrelaterade prov, i det senare om målrelatera-

⁶⁷ I allmänhet anges poängen på en skala med medelvärdet 0 och standardavvikelse 1. Denna kan dock transformeras till andra värden. Så transformerar t.ex. PISA, TIMSS och PIRLS skalan till medelvärdet 500 och standardavvikelsen 100.

Dnr 01-2003:2038

de eller kriterierelaterade. Det betygssystem som användes i Sverige mellan 1962 och 1994 var t.ex. grupprelaterat. Det viktiga var att rangordna eleverna på ett korrekt sätt, sedan fördelades betygen efter i förväg fastställda procent-satser, 7% ettor, 24% tvåor, 38 % treor, 24 % fyror och 7% femmor. Elevernas kunskaper jämfördes med andra elevers kunskaper. Den fördelning som redovisas ovan skulle dock gälla på nationell nivå och behövde inte alls vara giltig i en enskild klass eller skola. I en riktigt duktig klass kunde alla elever få betyget fem om de tillhörde de sju procent bästa i landet.

Ett normrelaterat prov har alltså en normgrupp mot vilken jämförelsen görs. I det tidigare svenska betygssystemet var normgruppen den grupp elever i Sverige som läste samma kurs samma år. Normgruppen för standardprovet i matematik 1975 var alltså alla elever som läste matematik årskurs 9 år 1975. Med en sådan normgrupp kan man inte avgöra om eleverna blir bättre eller sämre över tid eftersom normgruppen är "glidande".

Man kan också ha en "fix" normgrupp som fallet var med NAEP före 1992. Det innebär t.ex. att man bestämmer att den grupp som gjorde provet ett visst år är normgrupp och alla efterföljande årskurser jämförs med den gruppen (tills man bestämmer sig för en ny normgrupp). En sådan modell innebär dock problem med att jämföra olika prov med varandra och förutsätter att man på något tillförlitligt sätt kan skapa parallella prov (om man baserar proven på klassisk testteori). Fördelen är att man kan jämföra de olika årskullars resultat med varandra. Användningen av modern testteori minskar kraven på parallellitet och underlättar jämförelser över tid.

Exempel på normrelaterade prov med glidande norm är, som nämnts, de standardprov och centrala prov som användes i Sverige fram till 1994 där varje årskull jämfördes med sig själv. Dessa prov byggde på klassisk testteori och resultatjämförelser mellan olika år var inte möjliga.

Exempel på prov med valfri norm som syftar till jämförelse över tid är PISA och TIMSS. Dessa baseras på IRT som underlättar sådana jämförelser genom att resultat från olika provtillfällen kan läggas in i samma statistiska modell.

Det svenska högskoleprovet kan ses som en blandform. Det bygger på klassisk testteori och använder i huvudsak glidande normering. Genom att försöka skapa så parallella prov som möjligt vill man dock säkerställa att provresultaten blir jämförbara över tid.

Kriterierelaterade prov

Med ett kriterierelaterat provsystem menas ofta ett system där de visade kunskaperna relateras till en specificerad kunskapsdomän. Denna kunskapsdomän operationaliseras genom att ett stort antal uppgifter konstrueras för att "täcka in" den. Eleven bedöms sedan utifrån hur stor andel av kunskapsdomänen (uppgifterna) han eller hon behärskar. Bedömningen görs ofta i procent. Ett G skulle då kunna betyda att eleven behärskar 40% av kunskapsdomänen, ett VG att 70% behärskas osv. Ett sådant system ställer stora krav på definition av kunskapsdomänen och de uppgifter som definierar den. Denna form är förhållandevis vanlig i många länder (t.ex. USA) och tillämpas också ofta inom den

Dnr 01-2003:2038

svenska högskolan. (I det senare fallet kan man dock diskutera i vilken utsträckning kraven på strikt definition av kunskapsområden är uppfyllda.). Det är inte heller helt ovanligt att denna typ av prov förekommer inom den svenska skolan, möjligen som en följd av att lärarstuderande ofta blir prövade på det sättet under sin utbildning.

Med statistiska metoder kan man avgöra i vilken utsträckning och med vilken säkerhet de ”domänrelaterade” provresultaten kan generaliseras till hela kunskapsdomänen. I Sverige är det dock inte brukligt att provresultat redovisas med angivna felgränser.

Kriterierelaterade prov används ofta vid certifiering, dvs. då det gäller att avgöra om en person ska godkännas eller underkännas. Det svenska körkortsprovet är ett typexempel på ett sådant prov. I sådana fall är man inte intresserad av att skilja mellan duktiga provtagare långt ovanför godkändgränsen eller mindre duktiga långt under den. Man vill då framför allt ha goda provegenskaper (litet mätfel) i närheten av godkändgränsen. I denna typ av prov kan man också använda kriterier för underkänt (vilket vi inte har för betygen i skolan). Att köra mot rött ljus är t.ex. ett sådant ”underkäntkriterium” i körkortsprovet.

Det finns dock många typer av kriterierelaterade prov och många är inte baserade på hur stor andel av kunskapsdomänen som behärskas utan på verbalt formulerade kriterier eller standards.

Standardrelaterade prov

I samband med framväxten av standardtänkandet har man också börjat tala om standardrelaterade prov. De liknar de kriterierelaterade proven genom att de försöker ange vilka kunskaper eller förmågor en provdeltagare har, inte hur han eller hon förhåller sig i relation till andra provdeltagare. Det som utmärker ett standardrelaterat prov är att medan de mer traditionella kriterierelaterade proven antingen försöker avgöra om en person är godkänd eller underkänd, eller ange hur stor andel av en kunskapsdomän en person behärskar, så anger de standardrelaterade proven hur en person presterar i relation till verbalt formulerade standards.⁶⁸

Standardrelaterade prov konstrueras enligt Cizek (1998) främst för att matcha innehållsstandard (content standard), dvs. urvalet av uppgifter görs så att det så väl som möjligt följer den innehållsstandard som gäller. Grundat på bedömning av de valda uppgifterna i relation till gällande kvalitetskriterier (performance standards) fastställs sedan motsvarande poänggränser på provet. Detta kan göras på olika sätt varav några beskrivs i nästa avsnitt.

För Sveriges del gäller att vi i och med införandet av Lpo 94 lämnade det grupprelaterade betygssystemet. Det ersattes då med ett system som närmast liknar ett standardrelaterat system av det slag som beskrivs ovan. De bedömningsanvisningar som medföljer de nationella proven kan då ses som motsvarigheten till innehållsanpassade performance standards. I bedömningsanvis-

⁶⁸ Standardrelaterade prov ska inte sammanblandas med standardiserade prov. De senare syftar på att villkoren för proven och deras genomförande är standardiserade, dvs. sker under likartade förhållanden för alla t.ex. beträffande tid, hjälpmedel, vilken information som ges etc.

Dnr 01-2003:2038

ningarna diskuteras utförligt olika elevsvar i relation till kursmål och betygskriterier.

Avsaknaden av content standards i Sverige medför att det lokala valet av innehåll inte behöver sammanfalla särskilt väl med det innehåll som konstruktörerna av de nationella proven valt. De bedömningsanvisningar som åtföljer proven försöker beskriva hur de aktuella uppgifterna och olika svar på dessa kan relateras till kriterierna för olika betyg. Man kan alltså säga att de nationella proven är standardrelaterade i den meningen att de bygger på bedömning av svar i relation till betygskriterier (performance standards). De exemplifierar hur betygskriterierna kan tolkas i relation till ett visst givet innehåll. De är dock inte standardrelaterade i den meningen att de på ett tydligt sätt kan relateras till en i förväg given innehållsstandard eftersom en sådan i stor utsträckning saknas i de svenska kursplanerna. Men proven själva kommer genom sitt innehåll att med tiden definiera sin egen innehållsstandard, vilken sannolikt också kommer att uppfattas som en generell innehållsstandard för ämnet eller kursen där provet ges.

Standard setting

Det är en sak att säga att man vill ha prov som talar om vad elever kan i relation till vissa fastställda standards en annan att avgöra hur det ska gå till i praktiken. Ända sedan de kriterierelaterade proven infördes på 1960-talet har olika metoder för att fastlägga "cut scores" föreslagits.

Medan de normrelaterade proven har haft väl utvecklade statistiska och psykometriska metoder för att handha resultaten har hanterandet av resultaten på kriterierelaterade prov varit betydligt mer omdiskuterat. Detta har till största delen haft att göra med att det här handlar om bedömning av provresultat i relation till verbalt utformade utsagor om kunskapskvaliteter, dvs. standards. Det handlar inte om att tillämpa väletablerade statistiska tekniker för t.ex. indelning i percentiler, utan om att göra korrekta bedömningar av den samlade prestationen.⁶⁹

I grova drag kan man skilja på två olika tillvägagångssätt. Det ena innebär att man jämför de ingående provuppgifterna med standards och sedan aggregerar uppgiftsbedömningarna till en bedömning av hela provet. Det andra innebär att man bedömt elever utifrån standards (på basis av tidigare prestationer) och utgår från vad dessa elever presterat på provet. Man kan då t.ex. utgå från vad elever som bedömts ligga på Godkändnivå presterar, vad de som ligger på VG-nivå presterar osv. Man kan också kombinera de två förhållningssätten uppgiftsgranskning och elevbedömning, vilket kanske är det vanligaste.

En annan skiljelinje går mellan bedömning som baseras på enskilda uppgifter (atomistisk) och bedömning som baseras på hela provet (holistisk). Även i detta avseende är det vanligt att båda metoderna kombineras.

⁶⁹ Inslag av bedömning finns givetvis på uppgiftsnivå i alla former av prov (möjligen med undantag av flervalsfrågor, men där kommer bedömningen in mer vid valet av uppgifter). För en uppgift där en elev själv formulerar sitt svar måste alltid en bedömning göras om svaret är rätt eller fel (enpoängsuppgifter), eller av hur många poäng det är värt (flerpoängsuppgifter). Även om man har noggranna rättningsanvisningar kommer det alltid att finnas bedömningslägen när det ska avgöras hur många poäng ett visst svar är värt.

Dnr 01-2003:2038

Proven kan bestå av uppgifter av olika slag. Det kan vara många korta uppgifter av typen flervalsuppgifter eller uppgifter med kortsvar. Eller det kan vara ett fåtal öppna uppgifter där det finns utrymme för många olika typer av svar. I det avseende skiljer sig inte normrelaterade och kriterierelaterade prov åt. Man kan inte genom att titta på ett prov avgöra om det är av det ena eller andra slaget. Det som skiljer är framför allt hur uppgifterna väljs och hur resultaten hanteras.

Det vanligaste sättet att bedöma provresultatet är att tillämpa ett poängsystem, dvs. olika uppgifter och olika lösningar på dessa renderar olika poäng som sedan summeras till en summapoäng för hela provet. Om ett prov prövar olika mål kan man tänka sig att dela provet i flera delar efter vilka mål de prövar, och sedan räkna varje del för sig. Eventuellt kan man sätta en godkändgräns för varje sådan del av provet.⁷⁰ Problemet med detta är dock att ju färre uppgifter man har i ett prov desto mindre reliabel blir mätningen (jämför Spearman-Browns formel). Var och en av provdelarna blir alltså mindre reliabel än helheten.

Om man å andra sidan summerar samtliga uppgifter kan hög måluppfyllelse på ett delmål dölja bristande måluppfyllelse för ett annat mål. Eftersom ett prov inte kan vara hur långt och tidskrävande som helst måste man alltså välja mellan att pröva uppfyllelse av många delmål på ett otillförlitligt sätt, eller att räkna samtliga delmål aggregerat med risk för att missa bristande måluppfyllelse för något delmål. Det vanligaste är man väljer det sistnämnda alternativet och utgår från den samlade provpoängen.

Angoffs metod

Givet att man utgår från den totala provpoängen gäller det att fastlägga vid vilken poäng gränsen för olika standards (eller betyg) ska ligga. Bland de metoder som utgår från bedömning av enskilda uppgifter är olika varianter av den så kallade Angoffs metoder de vanligaste. De går i korta drag ut på att man utser en grupp bedömare som ska ta ställning till de uppgifter som ingår i provet. Dessa är valda på ett sådant sätt att de uppfyller content standards så väl som möjligt. Bedömarna får sedan granska de olika uppgifterna en efter en och bedöma hur stor sannolikheten är att en elev som t.ex. ligger på gränsen mellan underkänd och godkänd ska klara uppgiften.⁷¹

Alternativt kan man tänka sig att bedöma hur stor andel av t.ex. hundra elever på gränsen mellan underkänd och godkänd man tror skulle klara uppgiften. Resultatet bör bli detsamma oberoende av om man bedömer sannolikheten för

⁷⁰ Ett sådant system skulle vara i enlighet med grundskoleförordningens föreskrifter att alla mål ska vara nådda för att en elev ska bli godkänd, såvida inte särskilda skäl föreligger, t.ex. funktionshinder (Grundskoleförordningen 7 kap §§ 7-9). Även om det i provmaterialen betonas att proven inte kan pröva alla mål så skulle en strikt tillämpning ändå innebära att av de mål som prövas måste samtliga vara uppfyllda för att provbetyget ska bli Godkänd.

⁷¹ Här kunde man kanske tänka sig att alla som uppfyller ett mål kan lösa en uppgift som prövar det målet, dvs. att lösningsprocenten skulle vara 100. Så ser det emellertid aldrig ut i verkligheten. Mycket duktiga elever kan ibland svara fel på tämligen enkla uppgifter och en elev som inte bedöms som speciellt kunnig kan ibland klara av en förhållandevis krävande uppgift. Detta kan visa med s.k. itemkaraktäristiska kurvor (se t.ex. redovisning av resultat från nationella kursprov i matematik B-D).

Dnr 01-2003:2038

en enskild individ av viss typ eller relativa frekvensen för en grupp av motsvarande individer.

När alla bedömnarna gjort sina bedömningar sammanställs resultaten i medelvärden och spridning varefter deltagarna samlas och diskuterar sina bedömningar, såväl av de enskilda uppgifterna som av den sammanlagda provpoäng som blir resultatet. Ofta kan bedömningarna variera ganska mycket vid den första genomgången. Därefter görs samma procedur om en eller flera gånger. Avsikten är att gruppen så småningom ska bli någorlunda eniga om hur stora sannolikheterna är för att ”gränselever” ska lösa de olika uppgifterna, dvs. att interbedömarreliabiliteten ska bli så hög att den bedöms tillräcklig.

Summan av de sannolikheter man enats om för varje uppgift (multiplicerade med eventuell poängfaktor för uppgiften) anger sedan var på poängskalan gränsen för godkänd ligger (om sannolikheten gällde elever på gränsen mellan IG och G). Samma procedur kan sedan upprepas för eventuella andra gränser på provet.

Ibland kombineras ovanstående metod med att empiriska data från utprovningar tillhandahålls.

En metod som Angoffs, som bygger på en grupp personers mer eller mindre subjektiva bedömningar, innebär givetvis en viss osäkerhet. Denna osäkerhet, interbedömarreliabiliteten, bör anges när proven och provresultaten presenteras. Om möjlighet finns bör även intrabedömarreliabilitet anges. Den innebär att samma personer gör om sina skattningar med visst tidsintervall för att bedöma om den egna bedömningen är konsistent. Även de procedurer som använts för framtagandet av betygsgränser eller cut scores bör ingå i redovisningen av provresultat (t.ex. ETS, 2000, Standard 8.3 och 8.4).

Contrasting groups

Olika former av Angoffs metod har länge varit den vanligaste. De kan egentligen sägas utgöra en blandform av bedömning av individ och bedömning av uppgift. En mer renodlad individbedömning ligger bakom metoder av typen ”contrasting groups”. Då låter man en utprovningsgrupp genomföra provet varefter man jämför hur elever som av läraren bedömts som t.ex. Basic, Proficient och Advanced presterat på provet. Dessa gruppers poängfördelningar på provet ritas upp i ett diagram och vid den poäng där kurvan för Basic skär kurvan för Proficient sätts gränsen mellan dessa två standards (eller betyg). På samma sätt fastställs övriga gränser.

En likartad metod skulle t.ex. kunna tillämpas på vissa av de nationella proven. Man skulle då kunna gå tillväga så att när provet är genomfört rapporterar de lärare Skolverket valt ut resultaten för sina elever, samt sin egen bedömning av vilket betyg de olika eleverna är värda (utifrån en samlad bedömning av de tidigare prestationerna). Därefter kan Skolverket gå till väga på samma sätt som ovan. Poängfördelningen för elever som av läraren bedöms ligga på nivån att de ej uppnått målen (EUG) ritas, liksom kurvan för elever som bedöms ligga på G. Det värde på poängskalan där dessa kurvor skär varandra skulle då ange poänggränsen för betyget G. Sedan upprepas samma procedur för övriga betyg.

Dnr 01-2003:2038

En sådan metod skulle ha fördelen att det var lärarkårens egen tolkning av betygskriterierna och bedömning av eleverna som låg till grund för betygssättningen på de nationella proven. Nackdelar skulle kunna vara att man inte i förväg kan ange var betygsgränserna på provet ligger, och att konstruktörsgruppernas inflytande minskar. Man kan också rikta mer principiell kritik mot ett sådant förfarande, t.ex. när det gäller den direkta kopplingen mellan provresultat och mål. Men samtidigt vore det ett sätt att tydligt göra en större del av professionen delaktig i bedömningen av de nationella proven.

Bookmarking

En ytterligare metod för att bestämma gränser är den som kallas Bookmarking. Den metoden används ofta i samband med standard setting av prov som baseras på IRT (Kiplinger, 1997). Denna metod utgår också från bedömning av uppgifterna, men inte enskilt utan först sedan de grupperats efter svårighetsgrad.

När provet har genomförts analyseras resultaten med hjälp av någon IRT-modell. Därvid tilldelas varje provdeltagare ett mått på förmåga (en latent poäng) och varje uppgift en "svårighetsparameter". Båda måtten ligger på samma skala. Uppgifterna rangordnas efter svårighetsgrad och grupperas i olika "Booklets", till exempel den lättaste tredjedelen i ett häfte, den mellersta tredjedelen i ett andra och den svåraste tredjedelen i ett tredje. Därefter får en grupp granskare (15-20) uppdelade i mindre grupper bedöma de olika häften med uppgifter och försöka formulera verbala beskrivningar (standards) av vilka förmågor eller kompetenser som behövs för att lösa uppgifterna. När varje grupp gjort detta samlas man och diskuterar sina formuleringar varefter proceduren upprepas några gånger tills rimlig enighet nåtts. I det avseendet liknar tillvägagångssättet Angoffs metod.

Det som är slående med denna metod är dock att performance standards formuleras utifrån ett empiriskt underlag. Man konstruerar alltså först uppgifterna (utifrån fastlagda content standards) och låter därefter eleverna genomföra provet. När resultaten behandlats statistiskt rangordnas och grupperas uppgifterna efter hur svåra de varit för eleverna och därefter formuleras alltså standards efter hur uppgifterna i respektive grupp ser ut. Det är således en rent empiristisk metod för att fastställa standards. Den går också i omvänd ordning mot annan standard setting där man först formulerar standards och sedan försöker bestämma cut scores. I det här fallet går man alltså tvärtom och bestämmer först cut scores och formulerar därefter standards. Metoden är alltså egentligen inte en metod för standard setting utan en metod för att formulera standards när cut scores är satta.

Det är alltså uppgifternas empiriskt erhållna svårighetsgrad som avgör i vilken booklet de hamnar. Det är därmed viktigt att det förutom innehållsstandards finns noggranna anvisningar ("frameworks") och riktlinjer ("guidelines") för vilka förmågor proven avser att mäta och vilken typ av uppgifter som ska ingå. En uppgift ska vara svår för att den prövar något som är dels är väsentligt och dels kräver djupa kunskaper. Uppgiften ska inte vara svår för att den gäller något perifert som få känner till.

Dnr 01-2003:2038

Bookmarkingmetoden har sina rötter i NAEP och tillämpas i t.ex. PISA, PIRLS och TIMSS när man konstruerar olika kunskapsnivåer för exempelvis läsförståelse. Dessa instrument har också välutvecklade anvisningar och riktlinjer för studierna och tillhandahåller tydliga exempel på vilken typ av uppgifter som ingår och hur de är kopplade till de kompetenser som avses bli utvärderade.

Några ytterligare aspekter

Det finns givetvis ytterligare aspekter av betydelse när det gäller prov och andra instrument för kunskapsbedömning och kunskapsutvärdering. Viktiga tekniska och andra överväganden som behöver göras gäller t.ex. DIF (Differential Item Functioning) som gäller undersökningar av om enskilda uppgifter missgynnar någon grupp av provdeltagare, t.ex. elever med utländsk bakgrund, kvinnor eller män, eller någon annan grupp. Det är också viktigt att granska att uppgifterna inte innehåller stereotyper som kan uppfattas som rasistiska eller sexistiska. Vidare är det väsentligt att undersöka om proven som helhet är "biased", dvs. otillbörligt gynnar någon grupp, t.ex. genom att egenskaper som inte ingår i det som provet ska pröva påverkar resultatet.

Prov fungerar också olika beroende på om de är "high stake" eller "low stake". Det förra innebär att de är av stor betydelse för provtagaren i något avseende, det senare motsatsen. Högskoleprovet är ett svenskt exempel på ett high stake-prov eftersom det kan vara direkt avgörande för om en provtagare kommer in på den universitetsutbildning han eller hon eftersträvar. Diagnostiska prov kan ses som low stake eftersom de inte får några avgörande konsekvenser för provtagaren. Detsamma kan sägas om utvärderingsinstrument som PISA, TIMSS och NU, vilka inte heller har någon särskild betydelse för den enskilda eleven.

De nationella provens ställning är mer oklar. I grundskoleförordningen (7 kap. 10 §) sägs att de ska "användas i slutet av årskurs 9 för att bedöma elevernas kunskapsutveckling"⁷² och som stöd för betygssättning. Samtidigt sägs i § 5 i samma kapitel att "Betyg skall sättas av lärare." Vilken roll de nationella proven ska ha i sammanhanget är alltså inte särskilt klart uttryckt. I Skolverkets kommentarmaterial "Bedömning och betygssättning" och i informationen som medföljer de nationella proven betonas dock att proven inte ska ses som examsprov utan som ett betygsunderlag bland andra. Samtidigt betonas i det senare materialet hur viktig sekretessen är för att proven ska fylla sin funktion, vilket kan ses som en signal som brukar vara typisk för high stake-prov. Skolverkets olika undersökningar, t.ex. kvalitetsgranskningen av betygssättningen Skolverket (2000), tyder på att det råder stor osäkerhet bland lärarna om vilken roll de nationella proven har.

Madaus (1991) hävdar enligt Carr m.fl. (2001) att även om proven inte är av avgörande betydelse för den enskilda eleven, men kan vara det för läraren eller skolan - t.ex. genom att resultaten publiceras offentligt - så kan man tala om high stake prov. I det perspektivet blir nationella prov high stake i och med att

⁷² Man kan tycka att den aktuella formuleringen är något tveksam. För att bedöma *utveckling* är det nödvändigt att ha åtminstone två mätpunkter. Man kan inte mäta utveckling i en punkt, dvs. vid ett enda tillfälle. Ett nationellt prov ger alltså inte underlag för att bedöma kunskapsutveckling. Däremot kan man naturligtvis pröva måluppfyllelse vid en viss tidpunkt.

Dnr 01-2003:2038

resultaten offentliggörs på skolnivå (och än mer om de skulle offentliggöras på gruppnivå), vilket kommer att ske från och med 2003 när totalinsamling av provresultat inleds. Dessa kommer sedan att publiceras på skolnivå inom ramen för SIRIS⁷³ på Skolverkets hemsida.

Vad används provresultat till?

När det gäller de stora rikstäckande provprogrammen, vilka är de som främst är av intresse i det här sammanhanget, så syftar de i flertalet länder främst till att ge beslutsfattare, allmänhet, skolor m.fl. information om kunskapsläget i landet, kommunen, skolan eller klassen. Beroende på syftet kan provet baseras på stickprov eller genomföras av alla. Informationen gällde tidigare i huvudsak jämförelser mellan grupper av elever men efter standardtänkandets genombrott gäller informationen främst kunskaper och prestationer i relation till fastställda standards. Eventuellt kombinerat med att bedöma i vilken utsträckning olika mål nåtts. T.ex. att si och så många procent ska bli bedömda som Basic, Proficient eller Advanced.⁷⁴

Provresultaten används inte endast för information utan också för att ställa ansvariga till svars ("accountability" i engelskspråkig litteratur). Detta är vanligt i USA och där finns exempel på stater som t.ex. betalar bonus till lärare med elever som lyckas särskilt bra på NAEP eller andra statliga prov.

Jämförelser i standardbaserade system görs, som nämnts, främst mot standards, men också jämförelser över tid är vanliga, främst med syfte att se om olika utbildningsinsatser lett till förbättrade resultat.

Jämförelser mellan olika kommuner och skolor är också vanliga men kan vara missvisande som mått på kommunernas och skolornas insatser, eftersom det är välbelagt att skolframgång i hög grad är beroende av föräldrarnas socioekonomiska status. På vissa ställen försöker man därför konstruera prov eller bedömningar av provresultat som mer baseras på vad skolan tillfört ("value added"). Sådana program kan vara tekniskt komplicerade och förutsätter att man har tillförlitliga ingångsvärden, vilket inte alltid är fallet. Det kan också vara möjligt att med statistiska metoder kompensera för t.ex. socioekonomisk status på det sätt som görs i SALSA⁷⁵ (se t.ex. Saunders, 2000).

I ett internationellt perspektiv förefaller nationella provprogram med syfte att sätta betyg på enskilda elever tämligen ovanliga. Mer vanligt är examinerande prov där betyget ges av provresultet. Än vanligare är dock, i varje fall i de stan-

⁷³ Skolverkets Internetbaserade Resultat- och kvalitetsInformationsSystem.

⁷⁴ Här kan man för USA:s del notera en förskjutning i målsättning sedan 1980-talet. Då handlade det främst om "minimal competence testing", dvs. målen gällde främst att en viss andel skulle uppnå minimimålen. Detta visade sig dock få negativa konsekvenser för arbetet med de elever som inte låg i riskzonen genom att fokus ensidigt riktades mot endast en kategori av elever. Målen har därför ändrats så att man även anger hur stor andel som bör uppfylla standards över miniminivå. Detta kan jämföras med Sverige där det fortfarande är mycket vanligt i kommunernas skolplaner att det endast finns angivelser om hur stor andel av eleverna som ska nå kursmålen, dvs. bli godkända, eller som ska bli behöriga för gymnasieskolan. Däremot finns inga målsättningar när det gäller andelen elever med högre betyg.

⁷⁵ Skolverkets Arbetsverktyg för Lokala SambandsAnalyser finns tillgängligt på Skolverkets hemsida.

Dnr 01-2003:2038

dardbaserade systemen, att provens syfte är information, utvärdering, och att tjäna som underlag för att utkräva ansvar. Den svenska modellen med betygssstödjande funktion där inget provbetyg ges och där läraren själv bedömer vilken roll resultaten på nationella prov ska ha vid betygssättningen, förefaller unik. Antingen har man system där provet är direkt examinerande, eller så har man ett system där både ett provbetyg och lärarens betyg anges (t.ex. Danmark), eller som tredje alternativ har man nationella prov som är utvärderande (och eventuellt endast ges till stickprov av elever som i grundskolan i Finland) och inte är kopplade till betygssättningen.

En intressant och relevant fråga är om proven och provprogrammen kan fylla flera syften samtidigt. I det avseendet finns det olika uppfattningar hos olika forskare. Gilmore (2002) anger några forskare som anser att ett enskilt instrument inte ska ha mer än en funktion, men hävdar samtidigt att det är fler och fler som anser att provprogram måste kunna fylla flera syften. Som exempel anges att summativa prov även måste kunna fylla formativa syften. Även utvärderande prov som syftar till att bedöma om skolan fungerar tillfredsställande ses som en form av formativ utvärdering eftersom ändamålet med utvärderingen är att förbättra verksamheten. Det kanske främsta skälet är dock att om varje syfte ska uppfyllas med ett eget instrument blir bördan på elever och lärare för stor och alltför mycket tid tas från undervisningen till olika former av prövning. En risken är också att i en sådan konkurrenssituation får de för undervisningen viktiga diagnostiska och formativa instrumenten stå tillbaka för mer examinerande och summativa prov.

En annan risk enligt Gilmore är att om det finns nationella prov kan lärarna anse dessa tillräckliga och bemödar sig då inte om att utveckla den egna förmågan att bedöma elevernas kunskaper. Å andra sidan menar hon att välkonstruerade prov kan utgöra goda förebilder för lärares egen konstruktion av bedömningsinstrument. Det finns alltså som så ofta både möjligheter och risker. Vilket syfte som ska väga tyngst måste bedömas från fall till fall.

Är alla nöjda med tänkandet i standards?

Det är ingen tvekan om att "the standards movement" har haft ett kraftigt genomslag, framför allt i de anglosaxiska länderna. Flertalet länder och stater har i dag standards av olika slag, må de sedan kalla content eller performance standards, attainment targets, level descriptions eller något annat. Det gemensamma för flertalet är att de anger vad eleverna ska kunna efter genomgången utbildning, och att de anger vilka prestationer som krävs för att bli placerad på en viss nivå.

Standardtänkandet har införts med syftet att tydliggöra målen för utbildningarna och för att därigenom öka kvaliteten och effektiviteten i utbildningen. Det kan också ses som ett typiskt "top down" initiativ. Det är de centrala beslutsfattarna som satt målen och formulerat övergripande standards. Som ett led i att öka effektiviteten har dock utformningen av undervisningen och delvis tolkningen av standards ofta överlåtits till de lokala aktörerna. Ansvar för att målen uppnås utkrävs dock i regel från central nivå via olika former av utvärderingar eller inspektioner.

Dnr 01-2003:2038

Tänkande i angivna kunskaper i form av standards har lett till ett ökat intresse för att undersöka i vilken utsträckning elever - och därmed skolsystem - presterar i relation till uppställda mål och hur stor andel av eleverna som presterar på olika nivåer, samt till att jämföra resultatförändring över tid. Detta har i sin tur, främst i USA, lett till en ökad användning av såväl lokala som statliga prov.

Många har starka invändningar mot denna ökande provanvändning, inte minst pedagogiska forskare. Argumenten är flera. För att inte provkostnaderna ska bli orimliga krävs att instrumenten inte är alltför arbetskrävande,⁷⁶ vilket innebär att prov med maskinrättningsbara flervalsfrågor tenderar att bli dominerande. Kritiken mot den ökande användningen av standardiserade kunskapsprov (främst i form av flervalsfrågor) sammanfattas av Phelps (2003) i åtta punkter.

- Induce "teaching to the test" which, in turn, leads to artificial inflation of score;
- narrow the curriculum to a small domain of topics;
- tap only "lower order thinking" and hence discourage innovative curricula and teaching strategies;
- cause student achievement to decline;
- are unfair to minorities and women;
- are costly in terms of money and time;
- are overused in United States, especially in comparison with other countries; and
- are opposed by all those who truly care about children,

Alla punkter är givetvis inte giltiga för Sverige men flera argument framförs även hos oss.⁷⁷ Flervalsfrågor är t.ex. rätt ovanliga i de nationella proven. En viss risk för att "narrow the curriculum" kan finnas eftersom de nationella proven själva är de enda statliga dokument som anger ett sanktionerat konkret innehåll i ämnet. Av samma skäl finns också en risk för "teaching to the test" eftersom någon gemensam standard som kan ligga till underlag för både undervisning och provverksamhet saknas. Proven blir standard. Även återanvändning av samma uppgifter kan medföra vissa farhågor av samma slag eftersom lärarna kommer att känna till ett antal av de uppgifter som återkommer. Nya uppgifter till proven är ett sätt att undvika en medveten eller omedveten inskränkning till att undervisa mot eventuellt återkommande uppgifter.

Ytterligare kritiska synpunkter, som framför allt gäller den betydelse summativa prov har för elevers motivation, kan t.ex. återfinnas i en litteraturoversikt av Harlen och Deakin Crick (2002) och hos Shepard (1991), Black & Wiliam (1998), Popham (2002a, 2002b), Stiggins (2003). I flertalet fall uttrycks en oro för att proven i alltför hög grad blir "assessment of learning" istället för "assessment for learning".

⁷⁶ I USA och många andra länder rättas proven centralt. Man tilltror inte lärarna att själva rätta proven utan rekryterar särskilda rättare vilket naturligtvis höjer kostanden för proven väsentligt.

⁷⁷ Phelps är i själva verket förespråkare för provanvändningen och framför ett antal argument mot de kritiska punkterna.

Dnr 01-2003:2038

REFERENSER

- American Educational Research Association m.fl. (1999) *Standards for Educational and Psychological Testing*. Washington DC.
- Black P & Atkin M. (1997). Policy perils of international comparisons: The TIMSS case. *Phi Delta Kappa*, 79(1), 22-28.
- Black & Wiliam. (1998): *Inside the black box. Raising standards through classroom assessment*. <http://www.pdkintl.org/kappan/kbla9810.htm>
- Carr, M, m.fl. (2001). *The effects of curricula and assessment on pedagogical approaches on educationa outcomes*.
<http://www.minedu.govt.nz/index.cfm?layout=document&documentid=5610&indexid=1108&indexparentid=2107>
- Chatterji, M. (2002). Models and methods for examining standards-based reforms and accountability initiatives: Have the tools of inquiry answered pressing questions on improving schools? *Review of Educational Research Fall 2002*, Vol. 72, No 3.
- Cizek, G J. (1998) "Filling In the Blanks:Putting Standardized Tests to the Test," *Fordham Report*, Vol. 2, No. 11.
www.edexcellence.net/library/cizek.pdf
- CITO (2001). *Cito group, Netherlands National Institute for Educational Measurement*.
http://www.citogroep.nl/e_cens/our_international_projects/Capability_Statement_March
- Englund, T. (19986). *Curriculum as a political problem: Changing educational conceptions, with special reference to citizenship education*. Lund: Studentlitteratur.
- ETS (2000). *ETS standards for quality and fairness*. Educational testing service. Princeton, New Jersey. <http://www.ets.org/download.html>
- GAO (2003). *Characteristics of test will influence expences; information sharing may help states realize efficencies*.
www.gao.gov/cgi-bin/getrpt?GAO-03-389
- Gilmore, A. (2002). Large-scale assessment and teachers' assessment capacity: learning opportunities for teachers in the National Education Monitoring Project in New Zealand. *Assessment in Education, Vol 9, No 3*.
- Harlen, W, & Deakin Crick. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning.
http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/assessment/review_one.htm
- Kiplinger, V. L. (1997). *Standard-setting procedures for the specification of performance levels on standard-based assessment*.
www.cde.state.co.us/cdeassess/asperf.htm
- Linde, G. (2003). *Kunskap och betyg*. Lund: Studentlitteratur.

Dnr 01-2003:2038

- Lundgren, U, P. (1979). *Att organisera omvärlden. En introduktion i läroplans-teori*. Stockholm: Liber.
- Madaus, G. (1991). The effects of important tests on students. *Phi Delta Kappa*, 73(nov), 226-231.
- Marklund, S. (1987). *Skolsverige 1950-1975 Del 5: Läroplaner*. Stockholm: Utbildningsförlaget
- Messick, S. (1989). Validity. I R.L. Linn *Educational Measurment*, Third edition. New York: American Council on Education/Macmillan Publishing Co.
- Ministry of Education British Columbia (2002). Standards-based assessment within a standards-based education system.
http://www.bced.gov.bc.ca/achievement/litreview_bw.pdf
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M. & Flaherty, C. L. (2002). *PIRLS 2001 Encyclopedia*. <http://pirls.bc.edu>
- Phelps, P. P. (2003). *Why testing experts hate testing*.
www.fordhamfoundation.org/library/phelps.htm
- PISA (2000). *Measuring student knowledge and skills*. OECD
- Popham, J. (2002a). *Building test to support instruction and accountability*.
www.nea.org/accountability/buildingtests.html
- Popham, J. (2002b). *Implementing ESEA's testing provisions, Guidance from an independent comission's requirements*.
www.aasa.org/issues_and_insights/issues_dept/Commission_Report_Book.pdf
- Rebarber, T. & McFarland, T. W. (2002). Estimated cost of the testing requirements in the No Child Left Behind Act.
www.schoolreport.com/AWNCLBTestingCostsStudy.pdf
- Rychen, S. R. & Salganic, L. H. (2003). *Key Competences for a successful life and a well-functioning society*. Göttingen: Hogrefe & Huber Publishers.
- Vallberg Roth, A-C. (2001). Läroplaner för de yngre barnen. Utveckling från 1800-talets mitt till idag. *Pedagogisk forskning i Sverige*, årg 6 nr 4 s 241-269.
- SFSO (2003). *Contributions to the DeSeCo symposium*. Neuchatel: Swiss Federal Statistics Office
- Shepard, L. (1991). *Will national test improve student learning?*
http://www.cresst.org/products/reports_set.htm
- Sivesind, Kirsten; Bachmann, Kari; Afsar, Azita (2003) *Nordiske læreplaner*. Oslo: Læringssenteret
- Sivesind, K, Bachman, K & Afsar, A. (2003). *Nordiske læreplaner*, Oslo: Læringssenteret.
- Skolverket (2000). *Betygssättning, nationell kvalitetsgranskning*. Stockholm: Fritzes kundservice.

Dnr 01-2003:2038

- Skolverket (2000). *Grundskolan, kursplaner och betygskriterier*. Stockholm: Fritzes kundservice.
- Skolverket (2001). *Bedömning och betygssättning, kommentarer med frågor och svar*. Stockholm: Skolverket.
- SOU 1974:53. *Skolans arbetsmiljö*. Stockholm: Utbildningsdepartementet.
- Stern, P. (1999). *Effekter av ämnesprov i år nio*. Skolverket (dnr 97:2218).
- Stiggins (2003). *Balanced assessment: The key to accountability and improved student learning*. www.assessmentinst.com
- Sunders, L. (2000). *Accountability mechanisms and processes: "Value added: Telling the truth about school performance?"* National Foundation for Educational Research in England and Wales.
<http://www1.worldbank.org/education/est/resources/topic%20papers/Valueadded.doc>
- Tobiassen och Thomassen (2000). *Det nationella provsystemet – sett med norska ögon*. Rapport till Skolverket.
- Tuijnman A. C. & Postlewaite T. N. (1994). *Monitoring the standards of education*. Oxford: Elsevier Ltd.
- Utbildningsdepartementet. (2001). *Samverkande styrning. Om läroplanerna som styrinstrument*. Ds 2001:48. Stockholm: Fritzes offentliga publikationer.