



Att mäta kunskaper och skapa information

– en vägkarta för bättre uppföljning i skolsystemet

GABRIEL HELLER-SAHLGREN
Institutet för Näringslivsforskning



**Näringslivets
skolforum**
SWEDISH ENTERPRISE SCHOOL FORUM

Att mäta kunskaper och skapa information

Författare: Gabriel Heller-Sahlgren
Institutet för Näringslivsforskning

Maj 2024

Näringslivets skolforum, Stockholm

Omslag: Näringslivet skolforum

Innehållsförteckning

Sammanfattning	4
Introduktion	6
Kunskapsuppföljningar och information påverkar elevers utfall	9
Brister i det svenska systemet	12
Systemet för kunskapsuppföljningar	12
Systemet för att mäta skolkvalitet	14
Att mäta kunskaper och skapa information	16
Andra länders system	18
Danmark	18
England	20
USA	22
En vägkarta för bättre kunskapsuppföljning och information	24
Inför årliga adaptiva prov i grundskolan	24
Låt de (externt rättade) nationella provens utformning variera över tid och använd de adaptiva proven som nivåankare	26
Inför ett system för trendmätning på nationell nivå	28
Skapa relevant information på basis av kunskapskontrollerna	29
Inrätta ett bedömningscentrum med ansvar för all kunskapsuppföljning	30
Slutsats	31
Appendix	32
Referenser	35

Sammanfattning

Den här rapporten diskuterar hur man kan skapa ett bättre system för kunskapsuppföljning i den svenska grundskolan och hur resultaten i sin tur kan användas för att förbättra informationssystemet. Den drar lärdomar från forskningen om effekterna av kunskaps- och kvalitetsuppföljning samt hur andra länder – Danmark, England och USA – har utformat sina system dessa hänseenden.

Det finns forskningsstöd för att kunskapsuppföljningar och den information som skapas från dem påverkar elevers utfall positivt. Forskningen visar också att information om skolkvalitet ofta har positiva effekter på elevers resultat.

Tyvärr har det svenska systemet för kunskapsuppföljning historiskt sett inte varit tillräckligt bra. Detta delvis på grund av en stark decentralisering i vad som bedöms och hur det bedöms. Varken betygen eller de nationella proven har utformats för att kunna vara en tillförlitlig måttstock på elevers kunskaper (speciellt inte över tid).

Detta har också minskat värdet av den information som har kunnat extraheras från kunskapskontrollerna. Mer generellt har man aldrig ens försökt skapa mått på förädlingsvärden – resultaten justerade för elevers tidigare prestationer – för att mäta skolors bidrag till inläringen.

Likaså brister systemet för uppföljning av skolans utfall på nationell nivå, då den är begränsad till ämnen som inkluderas i internationella undersökningar såsom PISA. Dessa genomförs också numera med ganska många års mellanrum.

Det är avgörande att systemet som skapas är koherent och att dess olika delar samspelar på ett önskvärt sätt. För att förbättra kunskapsuppföljningen och informationssystemet i Sverige föreslås följande:

1. Inför årliga adaptiva prov i grundskolan

Man bör införa årliga självriktade prov i svenska, matematik, SO-ämnena och NO-ämnena från årskurs 1. Proven bör vara adaptiva, vilket innebär att uppgifternas svårighetsgrad anpassas till elevernas förmåga i realtid när de skriver provet. Inga krav bör finnas på att proven ska användas i (den snart reformerade) betygssättningen, men lärare ska kunna göra det på frivillig basis i den mån de själva ansvarar för den.

2. Låt de (externt rättade) nationella provens utformning variera över tid och använd de adaptiva proven som nivåankare

De (externt rättade) nationella provens utformning bör variera över tid. Exakt vilken specifikation som ges under vilket år kan slumpas ut strax innan proven skrivs. På detta sätt blir det svårare för aktörerna i skolsystemet att enbart rikta undervisningen mot ett specifikt slags prov. Resultaten på de adaptiva proven bör samtidigt användas som ankare för att säkerställa att kraven för olika resultat på de nationella proven är lika höga över tid.

3. Inför ett system för trendmätning på nationell nivå

Man bör skapa ett urvalsbaserat provsystem för att systematiskt studera hur kunskaperna förändras på nationell nivå. Prov i läsförståelse och matematik samt SO- och NO-ämnena bör skrivas vartannat år bland elever i årskurs 3, 6 och 9. Dessa prov bör utformas på ett liknande sätt som internationella undersökningar och bör inte ha några explicita incitament knutna till sig på lärar- eller skolnivå. Däremot kan man ge vissa incitament till elever för att öka sannolikheten att de anstränger sig.

4. Skapa relevant information på basis av kunskapsuppföljningarna

Individuella provresultat på de adaptiva proven och de nationella proven bör göras tillgängliga för rektorer, lärare, elever, föräldrar och myndigheter. Myndigheter kan sedan använda informationen för att skapa mått på skolors förädlingsvärden. Genom samspelet mellan de adaptiva respektive nationella proven begränsas, trots detta, skolornas incitament och möjligheter att enbart styra undervisningen i endera riktningen. Resultaten på de nationella trendmätningarna bör dock enbart publiceras och analyseras (förutom i avidentifierad form) på nationell nivå.

5. Inrätta ett bedömningscentrum med ansvar för all kunskapsuppföljning

Ett oberoende bedömningscentrum som är ansvarigt för all kunskapsuppföljning och analys bör inrättas. Detta kan med fördel förläggas vid ett universitet. Centrumet bör tillsätta en expertgrupp med uppdrag att analysera olika slags prov och informationsmått för att avgöra den exakta vägen framåt på basis av vägkartan i rapporten.

1. Introduktion

Under de senaste decennierna har kunskapskraven på den svenska arbetsmarknaden ökat markant, i takt med att teknologisk utveckling har rationaliserat bort många av de enkla jobb som tidigare existerade (Tåhlin 2007). Sammantaget har den ekonomiska avkastningen på både kognitiva och icke-kognitiva färdigheter därför ökat avsevärt i Sverige sedan tidigt 1990-tal (Edin m.fl. 2022). Över lag finns dessutom ett tydligt samband mellan elevers kunskaper och länders ekonomiska tillväxt (Heller-Sahlgren och Jordahl 2023). För att säkra framtidens kompetensförsörjning och tillväxt behövs därför ett större fokus på kunskaper i skolsystemet.

Ett skolsystem med fokus på kunskap kräver i sin tur rigorösa kunskapsmått och ett robust system för kvalitetsmätning. Information om elevers progression och skolors kvalitet är avhängiga de prov och kunskapskontroller som används för att följa upp elever, lärare och skolor. Och för att myndigheter och huvudmän ska kunna identifiera och åtgärda problem i skolsystemet måste de veta var problemen finns – och bland vilka elever – vilket också kräver en systematisk kunskapsuppföljning.

Den här rapporten diskuterar hur man kan skapa ett bättre system för kunskapsuppföljning i grundskolan – och hur resultaten från denna kan omvandlas till relevanta kvalitetsmått, för att förbättra informationssystemet på olika nivåer. Det finns forskningsstöd för att kunskapsuppföljningar och den information som skapas från dem påverkar elevers utfall positivt. Forskning visar också att information om skolkvalitet ofta har positiva effekter på elevers resultat.

Tyvärr har det svenska systemet för kunskapsuppföljning historiskt sett inte varit tillräckligt bra. Detta beror delvis på den starka decentraliseringen i provutformning och rättning, vilket har gjort det svårt att veta i vilken utsträckning elever ligger i fas med de kunskapskrav som ställs. Varken betygen eller de nationella proven har utformats för att kunna vara en tillförlitlig måttstock på elevers prestationer (speciellt inte över tid).

Detta har i sin tur minskat värdet av den information som har kunnat extraheras från kunskapskontrollerna. Mer generellt har man aldrig ens försökt skapa ett mått på skolors förädlingsvärde, som mäter skolors resultat efter justering för elevers tidigare utfall – något som forskning visar är avgörande för att man ska fånga skolornas bidrag till inläringen.

Likaså brister systemet för uppföljning av skolans utfall på nationell nivå. I dag används huvudsakligen internationella urvalsundersökningar, såsom PIRLS, PISA och TIMSS, för detta syfte. Dessa fyller en viktig funktion, men proven är begränsade till ett fåtal ämnen – och de genomförs med flera års mellanrum.

För att stärka det svenska kunskapsuppföljnings- och informationssystemet är det viktigt att studera vad vi kan lära från andra länder. Det finns flera viktiga lärdomar att hämta från systemen i Danmark, England och USA, vilket rapporten nyttjar för att skapa en vägkarta för Sverige.

Tyvärr har det svenska systemet för kunskapsuppföljning historiskt sett inte varit tillräckligt bra.

Det är avgörande att systemet som skapas är koherent och att dess olika delar samspelar på ett önskvärt sätt. Till att börja med bör man införa årliga läroplansbaserade självvärtade prov i grundskolan i svenska, matematik, SO-ämnen (ett prov) och NO-ämnen (ett prov). Proven bör skrivas under tidig hösttermin och vara adaptiva, vilket innebär att uppgifternas svårighetsgrad anpassas till elevernas förmåga i realtid. En fördel med dessa prov är att de kan vara ungefär 50 procent kortare än vanliga prov, men ändå estimeras låg- och högpresterande elevers kunskaper på ett mer tillförlitligt sätt. Inga krav bör finnas för att proven ska användas i betygssättningen, men lärare ska kunna göra det på frivillig basis i den mån de själva ansvarar för den.

Rapporten rekommenderar också att man låter de nationella provens utformning variera över tid, med de adaptiva proven som nivåankare. Just nu genomförs en utredning för att göra betygssättningen mer likvärdig, vilket sannolikt kommer att mynna ut i att den på något sätt kopplas till de (externt rättade) nationella proven. Detta kommer stärka incitamenten att styra undervisningen mot dessa prov markant.

Genom att ändra de nationella provens format från år till år blir det svårt för aktörerna i skolsystemet att enbart träna eleverna för att klara dem. Med resultatet på de adaptiva proven som ankare kan man samtidigt säkerställa att kraven för olika resultat på de nationella proven är lika höga över tid. Detta betyder att relationen mellan resultaten på de adaptiva proven och de nationella proven inte kan ändras över tid på nationell nivå.

Rapporten rekommenderar också att Sverige skapar ett urvalsbaserat provsystem för att systematiskt studera hur kunskaperna förändras på nationell nivå. Prov i läsförståelse och matematik samt SO- och NO-ämnen bör skrivas vartannat år bland elever i årskurs 3, 6 och 9. Resultaten bör inte ha några incitament knutna till sig på lärar- eller skolnivå. Däremot kan man ge vissa incitament till elever för att öka sannolikheten att de anstränger sig på proven.

Elevers individuella provresultat på de adaptiva proven bör göras tillgängliga för rektorer, lärare, elever, föräldrar och myndigheter. Myndigheter kan sedan använda informationen för att skapa förädlingsvärden på skolnivå. Förädlingsvärdena som skapas utifrån de adaptiva proven bör beräknas som genomsnittet av skolors förädlingsvärde, mätt över alla enskilda år.

Provresultaten på de nationella proven bör också göras tillgängliga på individ- och skolnivå på samma sätt som de adaptiva proven. Detta innebär att man skapar ytterligare ett förädlingsvärde på basis av kunskapsresultaten i de nationella proven. Man kan sedan skapa ett genomsnittligt förädlingsvärde utifrån det som skapas från de adaptiva proven och det som skapas utifrån de nationella proven.

De adaptiva och nationella proven – med olika format och användning – samspelar på ett sätt som gör att incitamenten och möjligheterna att styra undervisningen i endera riktningen begränsas, trots att man använder proven för att skapa information om skolors kvalitet. Fokuserar man alltför mycket på de adaptiva proven riskerar man att misslyckas på de nationella proven, vars format förändras över tid och är avgörande för betygssättningen (och förädlingsvärden); fokuserar man alltför mycket på de nationella proven riskerar man å andra sidan att misslyckas på de adaptiva proven, vilket i sin tur skulle påverka skolornas

förädlingsvärden negativt. Dessutom skulle det indirekt påverka möjligheterna att få höga resultat på de nationella proven, givet att de adaptiva proven fungerar som ett nationellt ankare för bedömningen. Över lag minskar således incitamenten och möjligheterna att styra undervisningen i endera riktningen.

Vad gäller resultaten på de nationella trendmätningarna bör dessa enbart publiceras och analyseras (förutom i avidentifierad form) på nationell nivå. Detta för att säkerställa att inga incitament knyts till kunskapsmätningarna på skol- och lärarnivå. Däremot kan man ge vissa incitament till elever för att öka sannolikheten att de anstränger sig.

Till sist föreslår rapporten att man inrättar ett oberoende bedömningscentrum som är ansvarigt för all kunskapsuppföljning och analys. Ett sådant centrum kan med fördel förläggas vid ett universitet och bör knyta till sig stark kompetens inom psykometri och nationalekonomi.¹

Ovanstående vägkarta är enbart en grov skiss. Innan man bestämmer sig för exakt hur proven och måtten ska utformas rekommenderar rapporten att bedömningscentrumet som inrättas tillsätter en expertgrupp med uppdrag att analysera olika slags prov och informationsmått för att avgöra den exakta vägen framåt.

1: Psykometri är en gren inom statistik och psykologi som fokuserar på utveckling av bedömningar och mätningar inom exempelvis utbildning.

2. Kunskapsuppföljningar och information påverkar elevers utfall

Ett robust system för kunskapsuppföljning är viktigt för alla aktörer inom utbildningssystemet. Direkta insatser för att stärka elevers resultat – i skolan eller i hemmet – kräver exempelvis att man har kunskapskontroller som på ett rättvisande sätt visar i vilken utsträckning elever når de kunskapskrav som ställs i läroplan och andra styrdokument. Likaså krävs ett robust system för kunskapsuppföljning för att man ska kunna veta hur elevers prestationer förändras över tid, exempelvis med anledning av sådana insatser.

Mycket riktigt tyder forskning på att kunskapsuppföljningar och den information som skapas från dem påverkar elevers utfall positivt. På ett allmänt plan tenderar prov att påverka elevers kunskaper positivt via kognitiva processer. Genom att man testas på saker som man har lärt sig plockar man fram kunskaper ur långtidsminnet, vilket gör att lär man sig mer. Mer frekventa kunskapskontroller gör att denna process måste upprepas, vilket gör att man lär sig ännu mer (se Jonsson m.fl. 2021; Roedinger m.fl. 2011). Frekventa kunskapskontroller gör också att elever börjar studera mer (och mer regelbundet), vilket i sin tur höjer deras prestationer (Roedinger m.fl. 2011). Det finns alltså generellt stöd för att prov och kunskapsbedömningar kan ha direkt positiv effekt på elevers utfall via kognitiva och beteenderelaterade mekanismer på elevnivå.

Men effekterna av att skriva prov är även indirekta via den feedback som elever får och reagerar på. Kunskapskontroller ger nämligen elever mer information om deras egna svagheter och styrkor, vilket gör att de kan lägga fokus på rätt saker i sina studier (Roedinger m.fl. 2011). Forskning finner exempelvis att lärares feedback om elevers (relativa) prestationer på tidigare prov – och deras kunskapsutveckling över tid – stärker elevernas resultat på nästkommande prov (se t.ex. Fischer och Wagner 2023; Hermes m.fl. 2021). Likaså finns det evidens för att elever presterar bättre när de, mer generellt, får information om hur de presterar jämfört med andra elever (Azmat och Ireberri 2010).² Att få mer information om sina egna prestationer på olika kunskapskontroller – och hur man presterar i relation till andra elever – kan alltså i sig stärka elevernas kunskaper.

Samtidigt finns det stöd för att även lärare justerar sitt beteende i positiv riktning efter att de får ny information från externa kunskapskontroller. Forskning från San Diego tyder exempelvis på att lärare justerade undervisningen på basis av resultaten på obligatoriska diagnostiska prov i matematik – och att införandet av dessa prov stärkte elevers matematikkunskaper ganska markant (Betts m.fl. 2017). Med ny information från standardiserade kunskapskontroller kan lärare på så sätt stärka undervisningen på ett sätt som leder till bättre utfall bland eleverna.

Likaså får föräldrar ofta information om elevers kunskaper med hjälp av resultaten på kunskapsbedömningar. Detta kan i sin tur hjälpa dem att justera sitt beteende på ett sätt

På ett allmänt plan tenderar prov att påverka elevers kunskaper positivt via kognitiva processer.

2: I svensk forskning finner man också att hur man rättar proven spelar roll för elevers utfall via direkta incitamenteffekter. I ett experiment i Stockholm fick sjätteklassare som skrev ett matematikprov som liknar det nationella provet information om hur de skulle bedömas precis innan provet började. Studien analyserar därför enbart rena effekter av att jobba hårdare på själva provet som ett resultat av bedömningssystemet. Man finner att relativ rättning har starkast positiva effekter på resultaten (Jalava m.fl. 2015).

som stärker elevernas utfall (se t.ex. Bergman 2019; Dizon-Ross 2019). Föräldrar i Los Angeles har exempelvis visat sig justera sin syn på sina barns ansträngningar och prestationer i skolan när de får mer information, vilket i sin tur gör att de inför striktare regler för eleverna i hemmet – inklusive dator- och mobilförbud – och elevernas prestationer ökade markant efter detta (Bergman 2021). Även i Australien, som har ett nationellt system för kunskapsuppföljning, finner man att föräldrar justerar sin syn på sina barns prestationer när de får jämförbar information. Efter att föräldrar får informationen får eleverna mer läxhjälp och gör mindre fritidsaktiviteter (Cobb-Clark m.fl. 2021). Det finns alltså starka tecken på att föräldrar reagerar på information om barnens prestationer på ett sätt som stärker fokuset på kunskaper.

Ovanstående tyder därför på att nationella system för kunskapsuppföljning har potential att stärka elevers prestationer. Ett exempel kommer från Danmark, där man införde (diagnostiska) digitala nationella prov år 2010.³ Det visar sig att elever som skriver proven i låg- och mellanstadiet presterar bättre i matematik och (framför allt) läsförståelse på grund av detta. Intressant nog är effekterna starkast bland elever med utländsk (icke-västerländsk) bakgrund och elever från lägre socioekonomisk bakgrund. Samtidigt visar det sig att skolor med en hög andel sådana elever är mer benägna att strunta i att anmäla sina elever till det obligatoriska provet (Andersen och Nielsen 2020).⁴ Mer generellt finns det också stöd för att olika former av standardiserade (nationella) prov har positiva effekter på elevers utfall i internationella undersökningar, både på kort och lång sikt (Bergbauer m.fl. 2024; Jürges och Schneider 2010; Leschnig m.fl. 2022). Över lag indikerar alltså forskningen att kunskapsutvecklingen stärks av prov inom ramen för nationell kunskapsbedömning.

Likaså är systemet för kunskapsuppföljning avgörande för att man ska kunna skapa relevant information med vilken man sedan kan identifiera effektiva och ineffektiva skolor. Utan sådan information kan myndigheterna inte stärka skolors incitament att förbättra sin kvalitet – och inte heller intervensera bland skolor som håller alltför låg kvalitet. Skolor och lärare kan då inte heller identifiera svagheter och styrkor inom den egna organisationen. Dessutom krävs ett sådant system för att föräldrar och elever ska kunna göra informerade skolval som grundas i kunskapsmässig kvalitet, då informationen de kan använda är avhängig systemet för kunskapsuppföljning (se t.ex. Heller-Sahlgren 2019; Heller-Sahlgren och Jordahl 2016).

Forskning visar mycket riktigt att information om skolkvalitet – och ansvarsutkrävande på basis av denna information – ofta har positiva effekter på skolors resultat (se Heller-Sahlgren 2019). I ett så kallat naturligt experiment i Storbritannien valde exempelvis Wales att sluta publicera information om skolors resultat samtidigt som England fortsatte att publicera denna information. Därefter försämrades provresultaten i grundskolan i Wales. Dessutom ökade skillnaderna mellan skolorna eftersom resultaten i de bästa walesiska skolorna inte försämrades (Burgess m.fl. 2013). Detta indikerar tydligt att olika aktörer inom skolsystemet reagerar på information om skolors resultat på ett sätt som stärker elevernas kunskaper.

3: Under det första året kraschade IT-systemet under två veckor, vilket gjorde att ett ganska stort antal elever plötsligt inte behövde skriva proven. Detta skapade slumpmässig variation i vilka elever som behövde skriva proven.

4: Annan forskning finner också att elever som underpresterar på matematikprovet i årskurs 3 höjer sig på provet i årskurs 6 som ett resultat av detta, vilket tyder på att elever, föräldrar och skolor reagerar positivt på informationen om elevers (under)prestationer (Beuchert m.fl. 2020).

Men de informationssystem som tycks fungera bäst bygger på så kallade förädlingsvärden: mått som isolerar skolors bidrag till inläringen, efter att man justerar för elevers bakgrund och, framför allt, tidigare resultat. Forskning visar att lärares och skolors förädlingsvärden kan mätas på ett sätt som fångar upp deras bidrag till elevers inläring och framtida utfall. System som mäter och offentliggör förädlingsvärden har ofta också positiva effekter på inläringen.⁵ Exakt hur bra systemen fungerar varierar beroende på hur man utformar dem, men även på kunskapskontrollerna och hur ofta dessa genomförs (se Britton m.fl. 2023; Heller-Sahlgren 2019; Heller-Sahlgren och Jordahl 2016).

Till sist är det också viktigt att notera att politiker och myndigheter behöver information om hur landets elevers kunskaper utvecklas i stort. Detta för att kunna avgöra hur väl skolsystemet fungerar – och för att kunna stärka det. Politiker i många länder har exempelvis reagerat kraftigt på undermåliga resultat i internationella undersökningar och försämringar över tid och påbörjat reformer som ett svar på detta (se Breakspear 2012). På grund av låga resultat i PISA svarade exempelvis tyska politiker med ett stort antal reformer, som också tycks ha haft positiva effekter på PISA-resultaten (Davoli och Entorf 2018). Likaså har sannolikt de reformer som Sverige genomförde runt 2010, bland annat som ett svar på fallande kunskaper i internationella undersökningar, bidragit till att höja elevernas prestationer (Heller-Sahlgren 2022). Med hjälp av information om hur landets elever presterar totalt sett kan man alltså (förhoppningsvis) ta bättre beslut för att stärka utfallen ytterligare.

Över lag tyder alltså forskningen på att kunskapskontroller är viktiga instrument för att stärka elevprestationerna, både via kognitiva och beteenderelaterade mekanismer hos olika aktörer i skolsystemet. Kunskapsbedömningar är generellt bra för inläringen och den information som aktörer får om individuella elevers prestationer höjer ofta kunskaperna i senare bedömningar. Likaså är kunskapsbedömningarna avgörande för att man ska kunna skapa rigorösa mått på skolkvalitet – och veta hur utbildningssystemet presterar över lag – vilket i sin tur är viktigt för olika aktörer i skolsystemet.

Forskning visar att lärares och skolors förädlingsvärden kan mätas på ett sätt som fångar upp deras bidrag till elevers inläring och framtida utfall.

5: Detta betyder naturligtvis inte att system som mäter förädlingsvärden är perfekta. Det finns saker som komplicerar utformningen och användningen av dem (se t.ex. Vlachos 2023), men dessa problem till trots tycks utbildningssystem fungera bättre med dem än utan dem (Heller-Sahlgren 2019).

3. Brister i det svenska systemet

Som föregående avsnitt visar tyder forskningen på att kunskapsuppföljningar och information från dem påverkar olika aktörer inom skolan på ett sätt som stärker elevers utfall. Det är därför ett problem att det nationella systemet för kunskaps- och kvalitetsuppföljning under lång tid har varit underutvecklat i den svenska grundskolan.

3.1. Systemet för kunskapsuppföljningar

Ett problem är den starka decentralisering av innehåll, kunskapskontroller och bedömning som präglar skolan. Betygssättningen har alltid skötts av lärarna och själva kunskapsinnehållet bestäms också till stor del av individuella lärare, med liten eller ingen extern kontroll kring vad som lärs ut. Historiskt har Sverige faktiskt aldrig haft ett system med någon form av extern bedömning i grundskolan (se Skolverket 2008). I stället har svenska lärare haft, och har, totalt ansvar för att bedöma och betygssätta sina egna elever, utan central moderering för att säkerställa att betygen är jämförbara (Wikström och Lind Pantzare 2018).⁶ Det visar sig föga förvånande att betygssättningen skiljer sig kraftigt mellan olika lärare och skolor, vilket framför allt tycks bero på "relativ betygssättning" i vilken bedömningen anpassas till skolornas genomsnittliga prestationsnivåer (se t.ex. Skolverket 2019).⁷ Alla kunskapskrav innefattar subjektiva tolkningar – och utan ett ankare för dessa tolkningar faller lärare tillbaka på relativ bedömning. Dessutom ges endast betyg från och med årskurs 6, vilket gör det omöjligt att använda dem för uppföljning innan högstadiet. Allt detta gör att betygen inte är en speciellt bra måttstock på elevers utfall i sig (eller progression över tid), varken inom eller mellan årskullar.

De enda enhetliga kunskapskontrollerna – där själva innehållet är precis likadant – har bestått av olika former av nationella prov. I och med att man införde efterkrigstidens relativa betygssystem skapades standardprov i olika årskurs (3, 6 och 8/9 efter att grundskolan infördes), med syfte att ge lärarna information om hur klassen över lag presterade i förhållande till andra klasser för att säkerställa betygens jämförbarhet. Eftersom så var fallet – och eftersom proven i praktiken syftade till att upprätthålla en förutbestämd nationell fördelning av betygen – var de inte heller utformade för att följa elevers kunskapsutveckling, även om man kunde följa elevers utveckling i betygsfördelningen som var tänkt att grundas i klassens prestationer på standardproven. Men detta var egentligen helt frivilligt: faktum är att proven inte var obligatoriska fram till ungefär ett decennium innan de skrevs för sista gången (och då endast i årskurs 9) och användningen av dem minskade dessutom över tid (se Ljung 2000).⁸

De enda enhetliga kunskapskontrollerna – där själva innehållet är precis likadant – har bestått av olika former av nationella prov.

6: Rapporteringen av betygen till SCB hanteras dessutom av de individuella skolenheterna och det sker ingen kontroll av att de uppgifter som rapporteras in till SCB är korrekta.

7: Trots den starka lokala autonomi kring kunskapsbedömningen och betygssättningen bygger samtidigt det centraliserade antagningssystemet till gymnasieskolan på att betygen som sätts av olika lärare och skolor är jämförbara, då de används som enda antagningsinstrument. Med några få undantag för estetiska och riksrekryterande utbildningar får inte gymnasieskolor anta elever på annat sätt än med meritvärdet från grundskolan (Skolverket 2024), som alltså beror på lärares egna bedömningar av sina elever.

8: Proven skrevs sedan grundskolan infördes i årskurs 3, 6 och 8/9. Sedan 1970-talet kom lärarna att minska användningen av standardproven – som alltså var helt frivilliga fram till slutet på 1980-talet – speciellt i tidigare årskurser, efter att betygen togs bort i låg- och mellanstadiet. Men även på högstadiet minskade användningen (Ljung 2000). Att standardproven inte var obligatoriska, men ändå utgjorde en så pass viktig pusselbit för betygssystemet att man starkt "rekommenderade" lärare att använda proven ledde till viss förvirring. En medlem i standardprovgruppen för matematik beskrev exempelvis i mitten på 1980-talet det politiska budskapet på följande sätt: "Standardprovet är frivilligt men det bör användas av alla" (Claesson 1986, s. 11).

Givet att de tekniskt sett var frivilliga är det kanske inte så konstigt att även standardproven under det relativa betygssystemet rättades av elevernas egna lärare (Skolverket 2008).⁹

När man under 1990-talet sedan övergav det relativa betygssystemet försvann standardproven och ersattes av nationella prov i engelska, matematik och svenska i årskurs 5 och 9. Endast proven i årskurs 9 var obligatoriska fram till och med 2008, även om de ändå användes i stor utsträckning. Resultaten i årskurs 5 samlades dock inte in systematiskt av myndigheterna innan 2009. De var också utformade enbart för att stämna av minimikrav snarare än hur elever presterar mer generellt sett (se Skolverket 2007, 2010a). Sedan 2010 skrivs obligatoriska prov i årskurs 3 i matematik och svenska, men även här ligger fokus på minimikrav snarare än mer allmänt på vilken absolut nivå elever presterar (se Heller-Sahlgren kommande). I årskurs 6 har dock proven fått samma karaktär som proven i årskurs 9 med probvetyg enligt skalan A-F. I årskurs 9 skriver även eleverna ett nationellt prov i ett av ämnena geografi, historia, religionskunskap eller samhällskunskap samt ett nationellt prov i ett av ämnena biologi, fysik eller kemi.¹⁰

Det finns indikationer på att utökningen av de nationella proven som skedde runt 2010 har haft positiva effekter på de svenska kunskapsresultaten, i enlighet med forskningen som diskuterades i föregående avsnitt (se Heller-Sahlgren 2022). Men de nationella proven är tyvärr inte speciellt bra för att följa upp elevers progression i sig, då de inte är utformade för detta syfte. Det går exempelvis inte att veta huruvida en elev som presterar sämre på de nationella proven i årskurs 9 än i årskurs 6 gör det på grund av att de faktiskt har försämrat sina kunskaper eller om kunskapskraven helt enkelt är lägre i årskurs 6. Likaså kan en sådan försämring bero på att proven har olika fokus i de olika årskurserna, vilket i praktiken är ett krav då antalet frågor som skulle krävas för att täcka hela kursplanens innehåll skulle bli för många.¹¹ Dessa problem – samt att läro- och kursplaner tenderar att reformeras över tid (utan att någon koppling görs i proven för att ta hänsyn till detta) – gör också att proven generellt inte lämpar sig för att jämföra resultaten mellan olika årskullar (se SOU 2014:12; SOU 2016:25).¹²

Att proven enbart skrivs vid tre tillfällen i grundskolan – och i årskurs 3 enbart med fokus på miniminivåer – är också ett problem för att man ska kunna använda proven för kunskapsuppföljning över tid.

Mer generellt är det också svårt för ett bedömningsinstrument att möta flera olika syften. Detta har varit fallet med de nationella proven, som historiskt sett har haft en mängd olika syften. Under lång tid var de huvudsakliga syftena att stödja en rättvis och likvärdig betygssättning samt fungera som underlag för måluppfyllelse på skol- huvudmanna- och nationell nivå. Men kraven för att proven ska fungera väl för dessa syften skiljer sig åt. Detta är också ett vanligt problem

Men de nationella proven är tyvärr inte speciellt bra för att följa upp elevers progression i sig, då de inte är utformade för detta syfte.

9: Utöver detta kan man notera att standardprovets riktmärke för betygssättningen luckrades upp något under 1980-talet. Enligt Lgr 1980 skulle standardproven användas som ett riktmärke för betyget 3. Eleverna skulle delas upp i tre grupper: de som får betyget 3, de som ligger lägre än betyget 3 och de som ligger högre än betyget 3. Provresultaten normerades alltså i praktiken efter en tregradig skala, trots att betygsskalan fortfarande var 1-5 (Ljung 2000).

10: Skolverket väljer slumpmässigt ut i vilket ämne respektive skolas elever ska skriva de två extra nationella proven. Under läsåret 2013/14 fanns det också obligatoriska prov i SO- och NO-ämnena i årskurs 6, men dessa gjordes frivilliga 2015 och togs sedan bort helt 2016. Liksom i årskurs 9 skrev eleverna endast ett (slumpmässigt utvalt) prov i SO-ämnena och ett prov i NO-ämnena.

11: Som framgår av diskussionen i Appendixet skulle detta problem minskas med hjälp av adaptiv provutformning. Men det går för tillfället inte att använda sådana prov på ett rättssäkert sätt om man inte enbart använder flervalfrågor eller frågor där enbart korta svar får anges.

12: Det går dock att använda de nationella proven för att studera hur elever rör sig upp och ner i fördelningen av provresultat och alltså hur deras kunskaper förändras relativt sett över tid, även om detta också kan påverkas av förändringar i provets innehåll. Om proven är helt relativa kan man dock inte mäta i vilken utsträckning elever når absoluta kunskapskrav eller förbättras absolut sett över tid.

i utformningen av kunskapsbedömningar: prov som fungerar väl för ett syfte fungerar inte nödvändigtvis bra för ett annat (se Skolverket 2017; SOU 2016:25). Det är därför viktigt att konkretisera ett primärt syfte för en given kunskapskontroll (Newton 2007) och Skolverket (2023) betonar numera enbart likvärdig betygssättning som huvudsakligt syfte för de nationella proven i årskurs 6 och 9. Detta efter en proposition som förtydligade detta och att proven ska ”särskilt beaktas” vid betygssättning. I årskurs 3 är syftet i stället att stödja bedömningen av elevers kunskaper i relation till kunskapskriterierna (Prop. 2017/18:14). Men proven i årskurs 3 fokuserar enbart på minimikrav. Att använda något av de nationella proven som huvudsakligt mått för att mäta elevers kunskapsutveckling över tid är alltså inte en bra idé.

Ovanstående problem blir ännu större när man betänker att de nationella proven rättas av elevernas egna lärare (precis som att betygen bestäms av elevernas egna lärare). Det sker ingen moderering av rättningen, vilket gör att provresultaten inte är jämförbara mellan olika elever. I analyser med omdöming av externa bedömare har man funnit att mer än hälften av elevlösningarna får olika bedömning av läraren och den externa bedömare (Skolinspektionen 2021). Det är också rimligt att även rättningen på de nationella proven påverkas av relativ bedömning, där lärarens rättning varierar beroende på elevernas genomsnittliga prestationsnivåer (Skolverket 2022). Eftersom bedömningen av de nationella proven varierar kraftigt mellan skolor minskar detta också utfallens användbarhet som mått på kunskaper över lag.

Utöver betygen och de nationella proven finns idag även internationella undersökningar, inklusive PISA, PIRLS och TIMSS, som används för att studera kunskapsutvecklingen i landet som helhet över tid och i jämförelse med andra länder. För detta syfte fungerar dessa prov väl. Samtidigt behandlar proven enbart ett begränsat antal ämnen och enbart specifika områden inom dessa ämnen (SOU 2016:25). Dessutom genomförs undersökningarna med ganska lång tids mellanrum. PISA har varit det prov som genomförts oftast (vart tredje år), men även PISA kommer från och med 2025 endast att genomföras vart fjärde år.

Det råder alltså ingen tvekan om att det nationella systemet för kunskapsuppföljning lider av brister. Varken betygen eller de nationella proven är utformade på ett sätt som gör dem till bra, jämförbara kunskapsmått på elev- eller gruppnivå – och speciellt inte för att mäta utvecklingen över tid. De internationella undersökningarna är värdefulla för att mäta kunskapsutvecklingen i det svenska skolsystemet som helhet, men även dessa prov lider av vissa brister.

3.2. Systemet för att mäta skolkvalitet

Likaså finns det brister i den information som skapas på basis av systemet för kunskapsuppföljning. Det finns inget system för att mäta skolors förädlingsvärden, i enlighet med den forskning som diskuteras i avsnitt 2. Skolverket har ett analysverktyg – SALSA – som justerar skolors genomsnittliga resultat i årskurs nio för (grova) bakgrundsvariabler. Inga indikatorer för elevernas tidigare resultat inkluderas i detta mått, vilket forskningen visar är avgörande för att förädlingsvärdesmått ska vara tillförlitliga. Dessutom skapas SALSA från data på skolnivå, vilket gör att man riskerar att justera för en del av den kvalitet som man försöker fånga upp (se Britton m.fl. 2023; Heller-Sahlgren 2019).

Det finns inget system för att mäta skolors förädlingsvärden.

En del av de problem som har identifierats med systemet för kunskapsuppföljning innebär naturligtvis också problem för möjligheterna att skapa bra mått på skolors förädlingsvärden; kunskapsmåttens utformning och syfte påverkar förädlingsvärdesmåttens kvalitet och precision. Eftersom det inte finns några nationella kunskapskontroller innan årskurs 3 kan man exempelvis inte följa upp lågstadieskolors kvalitet. Och eftersom de nationella proven i årskurs 3 fokuserar på minimikrav fungerar de sämre som indikatorer på elevers initiala förmåga i mellanstadiet jämfört med hur väl provresultaten i årskurs 6 fungerar som indikatorer på elevers initiala prestationer på högstadiet. Detta har noterats vara ett problem för att skapa förädlingsvärden för mellanstadiet (Holmlund m.fl. 2019).¹³

Samtidigt är vissa av problemen som har identifierats med kunskapsuppföljningen inte lika problematiska för syftet att skapa förädlingsvärden. Dessa är nämligen oftast relativa och mäter skolors prestationer i förhållande till varandra under ett specifikt år (se Heller-Sahlgren 2019; Holmlund m.fl. 2019). Förändringar i rangordningen över tid fångar därför förändringar i skolors relativa prestationer, vilket också minskar betydelsen av att kunskapskontrollerna är helt jämförbara över tid.

Det största problemet är dock den interna betygssättningen och rättningen av de nationella proven, vilket självfallet har implikationer även för måtten på skolkvalitet som skapas utifrån betygen och de nationella proven. Man har mycket riktigt varit rädd för att publicering av förädlingsvärden ska spä på betygsinflationen (Skolverket 2015). Detta stämmer sannolikt inte eftersom betygen och resultaten på de nationella proven redan idag används för att bedöma skolors utfall, exempelvis av Skolinspektionen. Frågan är inte heller om de mått som kan skapas skulle vara perfekta utan huruvida de vore bättre än att inte ha något sådant mått överhuvudtaget. Men det råder ingen tvekan om att systemet för betygssättning och rättning är ett problem för möjligheterna att skapa rigorösa mått på skolors kvalitet.

13: Det går dock att extrahera mer information på basis av de nationella provresultaten i årskurs 3 än vad som görs för tillfället (se Heller-Sahlgren kommande).

4. Att mäta kunskaper och skapa information

Som föregående avsnitt har visat vore det önskvärt att förbättra det svenska systemet för kunskapsuppföljning och den information som skapas utifrån detta. Här diskuteras översiktligt olika frågeställningar kring utformningen av både kunskapskontroller och förädlingsvärden.

Kunskapskontrollers allmänna relevans beror på deras validitet respektive reliabilitet. Med validitet menas att kunskapskontrollen ska fånga upp de saker man är ute efter att mäta, på ett sätt som gör att informationen som skapas utifrån den blir bra. Validiteten är även kopplad till hur man använder resultaten på kunskapskontrollen, då felaktig användning kan underminera validiteten. Hot mot validiteten handlar förutom detta huvudsakligen om att man mäter annat än det man har tänkt och att man täcker för lite av detta i kunskapskontrollen. Reliabilitet innebär i stället att kunskapskontrollen har hög mätsäkerhet, vilket kräver ordentlig utprovning av uppgifterna och att man har tillräckligt många uppgifter för att undvika att slumpmässiga faktorer avgör resultaten (se SOU 2016:25).

Som noterades i föregående avsnitt finns det problem med att ge kunskapskontroller alltför många syften, då det inte nödvändigtvis går att uppnå hög validitet inom alla syften. Varje användning av resultaten måste därför prövas var för sig och man bör enbart ha ett primärt användningsområde för en given kunskapskontroll (Newton 2007; Skolverket 2017; SOU 2014:12; SOU 2016:25).

De nationella provens primära syfte är idag exempelvis att säkerställa en likvärdig betygssättning i årskurs 6 och 9. Men givet att proven rättas internt och likvärdigheten i betygssättningen är låg kan man säga att de har låg validitet för detta ändamål. De har också låg validitet vad gäller det tidigare andra huvudsakliga syftet (som fortfarande är huvudsyfte för proven i årskurs 3) – att fungera som underlag för måluppfyllelse på skol-, huvudmanna- och nationell nivå – då de inte har fungerat väl heller för detta ändamål.

När man föreslår nya kunskapskontroller eller ändringar till existerande kunskapskontroller är det därför viktigt att man preciserar vilket huvudsakligt syfte de ämnar få. Lika viktigt är att man undviker att ge dem sekundära syften som går emot det primära syftet – eller syften som inte kan uppnås på grund av det primära syftet man har valt. Om man ämnar använda kunskapskontrollen för flera syften är det således viktigt att man noga tänker igenom utformningen av både proven och den information som skapas utifrån dem.

Ett exempel är att kunskapskontroller till vilka man knyter starka incitament för att utkräva ansvar ofta ger upphov till olika försök till manipulation, vilket riskerar att minska deras validitet för andra syften. Att lärare och skolor endast fokuserar på det som fångas upp i prov – eller i de kunskapskontroller som används – är ett vanligt argument mot att man använder standardiserade prov överhuvudtaget. Om man använder kunskapskontroller för att styra betygssättningen finns det samtidigt starka incitament att helt enkelt börja träna elever för att klara just dessa prov i stället för att lära ut bredare

Man bör enbart ha ett primärt användningsområde för en given kunskapskontroll.

kunskaper. Eftersom prov inte kan vara hur långa som helst och eftersom man ofta vill fånga upp elevers kunskaper inom olika områden i ett givet ämne (t.ex. algebra inom matematik) utan stora mätfel i respektive område riskerar detta leda till att undervisningen snuttifieras mot just de delar som ingår i provet (se Heller-Sahlgren 2019; SOU 2016:25). I utformningen av systemet för kunskapsuppföljning och informationen som baseras på detta bör man ta hänsyn till sådana och andra liknande problem för att säkerställa hög validitet för de syften man vill uppnå.

Likaså måste man säkerställa att de mått på skolkvalitet som skapas är valida. Som noteras i föregående avsnitt är forskningen tydlig med att det bästa måttet för att fånga upp skolors kvalitet är det så kallade förädlingsvärdet. Metodologiskt sett är det absolut viktigast att man har tillgång till tidigare resultat för att sådana mått ska fungera väl (se Britton m.fl. 2023; Heller-Sahlgren 2019). Med rätt data är den tekniska konstruktionen av förädlingsvärdena relativt enkel och redovisas i Appendixet. Det är dock avgörande att beräkningarna av förädlingsvärdena i sig inte underminerar de data som de baseras på. Detta skulle kunna vara fallet om det exempelvis leder till att lärarna börjar träna elever för att klara den specifika kunskapskontrollen som förädlingsvärdena skapas utifrån. Om så är fallet riskerar självfallet även måtten på skolors förädlingsvärden att undermineras.

I utformningen av både kunskapskontroller och den information som skapas utifrån dessa kunskapskontroller vill man helt enkelt minimera (och helst neutralisera) de problem som kan uppkomma till följd av att man använder kunskapskontroller för olika syften, enligt ovanstående diskussion. Målet är att skapa ett så sammanhängande system som möjligt totalt sett – och för detta ändamål är det av värde att först exemplifiera hur andra länder har valt att konstruera system för kunskapsuppföljning och information i skolan.

Att lärare och skolor endast fokuserar på det som fångas upp i prov är ett vanligt argument mot att man använder standardiserade prov överhuvudtaget.

5. Andra länders system

Det här avsnittet diskuterar hur man i tre andra länder – Danmark, England och USA – har skapat nationella system för kunskapsuppföljning och resultatinformation. Detta för att dra slutsatser kring hur Sverige kan skapa ett bättre system med hjälp av lärdomar från dessa länder.

5.1. Danmark

Det danska systemet för kunskapsuppföljning i grundskolan har förändrats under de senaste decennierna. Fram till och med 2010 utvärderades inte danska elevers resultat förrän i årskurs 8, men efter undermåliga resultat i PISA 2000 och 2003 infördes obligatoriska nationella prov i flera årskurser i alla kommunala skolor under läsåret 2009/10 – efter rekommendationer av både nationella utredningar och OECD – för att skapa bättre verktyg för kunskapsuppföljning och feedback.¹⁴ Dessa prov har sedan starten varit självvärtande och digitala samt har varit utformade för att man ska kunna följa elevers prestationer över tid (se Andersen och Nielsen 2020; Beuchert och Nandrup 2018; UVM 2023a).

De nationella proven var fram till och med läsåret 2021/22 adaptiva – vilket innebär att provfrågornas svårighetsgrad anpassas till elevernas förmåga i realtid när eleverna skriver provet (se avsnitt 6.1 och Appendix) – och skrevs under vårterminen vid upprepade tillfällen i läsning (årskurs 2, 4, 6 och 8), matematik (årskurs 3, 6 och, sedan läsåret 2017/18, även 8) och engelska (årskurs 7 och, sedan 2017/18, även 4).¹⁵ Dessutom existerade prov i fysik, biologi och geografi i årskurs 8, även om proven i biologi och geografi gjordes frivilliga från och med läsåret 2017/18. Dessutom har det funnits prov i danska som andraspråk i årskurs 5 och 7, som alltid har varit frivilliga att skriva (Beuchert och Nandrup 2018; UVM 2023a).¹⁶

Resultaten på proven har dels presenterats enligt en relativ skala efter hur eleverna presterade i ett pilotprov år 2010: (1) väldigt långt under genomsnittet, (2) under genomsnittet, (3) genomsnittliga, (4) över genomsnittet och (5) långt över genomsnitt. Genom att man förankrade resultaten till resultaten på pilotprovet ämnade man också använda resultaten för att studera förändringar i kunskaperna mellan årskullar över tid för att på det sättet utvärdera förändringar på systemnivå.¹⁷ I läsning och matematik har man även presenterat resultaten på en målrelaterad skala med sex nivåer: "Otillräcklig prestation", "Bristfällig prestation", "Jämn prestation", "God prestation" och "Utmärkt prestation" (Andersen och Nielsen 2020; Beuchert och Nandrup 2018; UVM 2023a).

Medan fokus har legat på elevers prestationer i relation till en nationell norm har man alltså kunnat följa elevers absoluta kunskapsutveckling i enlighet med

14: Proven har inte varit obligatoriska i danska friskolor.

15: Ett annat exempel på adaptiva provsystem av denna typ är Australiens NAPLAN-system (se NAP 2024).

16: Utöver detta har proven också kunnat skrivas två gånger per elev på frivillig basis under höstterminen – antingen under samma läsår eller under läsåret innan eller efter som de obligatoriska proven skrivs – men eftersom proven har varit adaptiva, i enligheten med diskussionen i avsnitt 6.1 och Appendix, har det inneburit att eleverna i praktiken då svarar på andra frågor.

17: Man standardiserade dock om de normrelaterade resultaten på basis av utfallen i årskullen som skrev proven under läsåret 2018/19.

nationella krav i läsning mellan årskurs 2, 4, 6 och 8 samt i matematik mellan årskurs 3 och 6 (och sedan läsåret 2017/18 även upp till årskurs 8).

Det nationella provsystemet genomgår dock just nu förändringar, som kommer vara färdiga från och med läsåret 2026/27. De adaptiva proven görs om till vanliga linjära prov, bland annat på grund av att man ville göra det lättare att förstå proven/resultaten samt öka transparensen kring provuppgifterna (för att på så sätt stärka möjligheterna att identifiera specifika svagheter och styrkor samt öka användningen av proven i skolan). Likaså kritiserades proven för att de var för svåra. Proven kommer enbart att vara obligatoriska i läsning och matematik, men antalet prov i dessa ämnen kommer att öka. I läsning kommer proven skrivas i årskurs 2, 3, 4, 6 och 8; i matematik kommer proven skrivas i årskurs 2, 4, 6, 7 och 8. Därutöver kommer det finnas frivilliga prov i fler ämnen, inklusive moderna språk. Proven kommer även i fortsättningen vara digitala och självvärtande, men kommer att förläggas i början av läsåret för att öka användningen av dem i planeringen av undervisningen samt minska sannolikheten att undervisningen enbart fokuserar på proven. Dessutom introduceras obligatoriska diagnostiska prov i årskurs 1 för att identifiera (1) elever med lässvårigheter och (2) begåvade elever samt en obligatorisk diagnostisk språkbedömning i förskoleklass (Eurydice 2024; Folkeskoleforligskredsen 2021).¹⁸

Däremot har Danmark inte något renodlat system som ämnar följa kunskapsutvecklingen på nationell nivå. I stället har man tänkt sig att de nationella proven ska kunna fylla detta syfte, samtidigt som deras primära syfte har varit att fungera som ett pedagogiskt verktyg för lärare och som utgångspunkt för skolors kvalitetsarbete (SOU 2014:12). Men givet att proven nu görs om helt har detta uppenbarligen inte varit en tillfredställande lösning, då det kommer att göra att man inte längre kan jämföra resultaten mellan årskullar före och efter förändringarna. Det hade sannolikt gått att länka proven trots förändringarna om man så hade velat, men givet deras huvudsyfte är det kanske inte så konstigt att man inte valde att göra det då en sådan länkning kräver ett mer komplicerat förförande.¹⁹

Utöver systemet för nationella prov har Danmark ett separat system med sju nationella, externt rättade examensprov i slutet på grundskolan (årskurs 9 eller årskurs 10), som är obligatoriska för elever i kommunala skolor. Målet är att andelen elever som får ett visst betyg motsvarar en förutbestämd nationell fördelning, men samtidigt ska varje enskild elev bedömas efter en målrelaterad skala. För att uppnå bägge utfall skapas proven med hänsyn till kopplingen mellan den önskade betygsfördelningen och de målrelaterade betygskriterierna – men i praktiken är således betygssystemet relativt på nationell nivå (Allerup och Kjeldsen 2017; UVM 2023b, 2023c).²⁰

Medan Danmark har ett förhållandevis välutvecklat system för kunskapsuppföljning i sig – även om man saknar ett separat system för att följa upp kunskapsutvecklingen i landet som helhet – är informationen om skolkvalitet inte lika utvecklad. Elever och deras föräldrar samt lärare får tillgång till resultaten på de nationella proven, medan rektorer får information om sko-

18: Bland eleverna som visar sig ha lässvårigheter kommer även en screening för dyslexi att bli obligatorisk.

19: Med nationella trendmätningar kan man dessutom ha fler uppgifter som täcker större delar av ämnesområdena; eftersom man inte ämnar fånga upp individuella elevers resultat, utan enbart resultaten för en given årskull, kan man låta elever svara på enbart ett urval av uppgifterna. Likaså finns det ett värde i att det inte finns några incitament alls för skolorna att prestera bra på proven för trendmätning, eftersom man då inte heller har incitament att försöka manipulera resultaten på olika sätt.

20: Precis som de nationella proven är examensproven inte obligatoriska för elever i friskolor, men eleverna måste skriva proven om de ska läsa vidare på en gymnasieutbildning som kräver att eleven har ett genomsnittligt betyg med sig från grundskolan (UVM 2023d).

lans genomsnittliga resultat på skol- och klassnivå. På skolnivå får de även informationen justerad för elevsammansättning, men inga föräldlingsvärden som tar hänsyn till elevernas tidigare resultat. Kommunerna har i sin tur tillgång till resultaten på skolnivå för de skolor som de driver. På nationell nivå har resultaten publicerats för allmänheten, men det är förbjudet att publicera enskilda skolors genomsnittliga resultat. En nationell myndighet har även tillgång till utfallen och kan använda dem för att följa upp skolor som underpresterar. Samtidigt är det ett krav att publicera skolors genomsnittliga resultat på examensskrivningarna (Nusche m.fl. 2016).²¹

Över lag har Danmark i flera hänseenden alltså under det senaste decenniet haft ett bättre system för kunskapsuppföljning än Sverige. Detta speciellt med tanke på systemet för nationella prov som har existerat sedan 2010. Även om systemet nu reformeras och proven kommer att ges i färre ämnen ska antalet prov i läsning och matematik samtidigt öka. Dessa digitala prov kommer att fortsätta att automaträttas och ge information till myndigheter, skolor och föräldrar om hur eleverna presterar över tid. Likaså finns ett examinations-system med extern rättning. Däremot saknar landet ett rigoröst system för att mäta kunskapsutvecklingen på nationell nivå och för utvärdering av skolors kvalitet på basis av de kunskapskontroller som existerar.

5.2.England

Det engelska systemet för kunskapsuppföljning bygger på baslinjetest, nationella prov och examinationsskrivningar som genomförs i slutet på grundskolan. Grunden för systemet lades 1988 då man introducerade en nationell läroplan för första gången och skapade så kallade "Key Stages" i utbildningen upp till och med slutet på grundskolan (och sedermera även gymnasiet). De nationella proven har historiskt sett skrivits i motsvarande årskurs 2 (Key Stage 1), 6 (Key Stage 2) och 9 (Key Stage 3), men numera är de endast obligatoriska i årskurs 6 (och är där externt rättade). Det har även funnits lärarbedömda kunskapskontroller – som modereras med externa censorer – i årskurs 2 och 6 i exempelvis naturvetenskap och skrivning. Därutöver skriver även elever externt rättade examensprov i cirka 10 ämnen i slutet på årskurs 11.

Det finns flera syften med de nationella proven i grundskolan. Dels ska de fungera som en kunskapskontroll av i vilken utsträckning elever når kunskapskraven i läroplanen. Man har också satt upp mål i detta hänseende och proven är tänkta att mäta i vilken utsträckning man når dessa mål. Proven används även för att mäta hur skolor presterar och för att utkräva ansvar i detta hänseende (se t.ex. STA 2023a). Exempelvis är resultaten viktig input till den engelska skolinspektionens arbete (se Ofsted 2024) och resultaten på proven i årskurs 6 publiceras offentligt. Examensskrivningarna i årskurs 11 används på samma sätt, men där avgör de även elevers slutgiltiga betyg i grundskolan och fungerar därför också som antagningsinstrument för vidare antagning (både till gymnasium och högskola).

Utöver de nationella proven har man nyligen infört en lärarrättad baslinjebedömning i engelska och matematik som görs under de första veckorna i

Över lag har Danmark i flera hänseenden alltså under det senaste decenniet haft ett bättre system för kunskapsuppföljning än Sverige.

²¹: Den huvudsakliga anledningen till att det är förbjudet att publicera danska skolors genomsnittliga resultat på de nationella proven är att man har velat skydda dem som ett pedagogiskt instrument och för att kunna följa kunskapsutvecklingen i landet som helhet. Som noterades i avsnitt 3 och 4 är det vanskligt att använda samma kunskapskontroll för en rad olika syften. Trots detta har lärare och rektorer länge antagit att skolor huvudsakligen hålls ansvariga på basis av dessa prov (Nusche m.fl. 2016). Som noteras ovan ville man minska sannolikheten för att proven används för detta syfte i och med de reformer som nu genomförs.

förskoleklassen. Bedömningen ersatte de externt rättade proven som skrevs i årskurs 2. Syftet med bedömningen är huvudsakligen att fungera som ett nolläge från vilket förädlingsvärden mellan förskoleklass och årskurs 6 kan skapas (STA 2020a). Dessutom finns det ett obligatoriskt test i avkodning ("phonics") som skrivs i årskurs 1 för att säkerställa att elever har lärt sig avkodning på en nivå som förväntas givet deras ålder samt ett obligatoriskt prov i multiplikation i årskurs 4 (STA 2017, 2023b).

Ett problem med provsystemet var länge att utformningen av examensproven inte fungerade för de syften man hade tilldelat dem. Trots att dessa rättas externt rådde mellan 1992 och 2010 betygsinflation som gjorde att proven i praktiken inte gick att använda för att mäta hur väl eleverna nådde kunskapskraven (Coe 2013). Det är generellt svårt att skapa målrelaterade prov som är lika svåra från år till år – och med starka incitament finns det en dessutom en risk att lärare börjar träna elever för att klara proven i stället för att lära ut alla slags kunskaper som läroplanen ska fånga upp. Endast när man började kräva att examinationsorganisationerna skulle använda tidigare resultat på de nationella proven i årskurs 6 för att statistiskt kalibrera betygssättningen fick man bukt på betygsinflationen i årskurs 11 (se Benton och Sutch 2014; Bramley och Vidal Rodeiro 2014; Ofqual 2018).²² Kalibreringen innebär i praktiken att betygsfördelningen justeras på basis av den statistiska relationen mellan resultaten på proven i årskurs 6 och utfallen på examensskrivningarna i en tidigare referensårskull; om resultaten på proven i årskurs 6 har ökat/sjunkit över tid så ökar/sjunkit även betygen på examensskrivningarna.²³

En intressant utveckling är att man sedan 2019 – med inspiration från Hong Kong – också använder ett nationellt referensprov i engelska och matematik i kalibreringssyfte, vilket skrivs av ett representativt urval av elever i årskurs 11. Frågorna är huvudsakligen desamma varje år, vilket gör att man kan jämföra resultaten över tid mellan årskullar. Resultaten på proven används sedan för att stärka kalibreringen av betygssättningen på examensskrivningarna i specifikt engelska och matematik (Ofqual 2023).²⁴ Samtidigt kan proven också användas för att fånga upp hur väl elevers kunskaper på nationell nivå har utvecklats över tid i dessa ämnen, på ett liknande sätt som i internationella undersökningar.

Man har också vidtagit åtgärder för att säkerställa att betygssättningen av de nationella proven i årskurs 6 är likvärdig sedan 2016, då man bestämde vilka kunskapsnivåer som krävs för olika resultat. Ett representativt urval av eleverna som skriver det riktiga provet skriver varje år också ett så kallat "ankarprov" som användes i provutformningen i 2016 års omgång. Man justerar sedan elevers provresultat på basis av jämförelser av urvalets prestationer på ankarprovet och det riktiga provet. På detta sätt kan man säkerställa att samma betyg motsvarar samma prestationer över tid och att endast förändringar i årskullens faktiska kunskaper leder till förändringar i provresultaten (se STA 2020b).

En intressant utveckling är att man sedan 2019 – med inspiration från Hong Kong – också använder ett nationellt referensprov i engelska och matematik i kalibreringssyfte.

22: Sådan kalibrering har visserligen använts som grund för betygssättningen under tidigare år, men då var systemet inte kodifierat – och betygsgränserna tenderade då att, om något, justeras uppåt jämfört med vad kalibreringen indikerade (se Benton 2016; Stringer 2012). I det nya systemet måste examinatorerna i praktiken rättfärdiga eventuella avvikelser från de statistiskt sett förväntade betygen, och myndigheten Ofqual måste godkänna den evidens som som godtagbar, för att betygen realistiskt sett ska kunna avvika från dessa (Beach 2015; Newton 2023).

23: Ett sådant system underlättar också förändringar i betygssystemet och kunskapskraven över tid, då man använder resultaten på de tidigare proven som ankare över tid (se Newton 2023). Då sådana förändringar sker ganska ofta är detta en styrka, jämfört med exempelvis i Sverige där övergångar till olika betygssystem har gjort att utfallen i praktiken inte går att jämföra före och efter förändringarna.

24: Detta kräver strikt sekretess och provförhållanden – och skrivningarna övervakas av externa monitörer från provkonstruktören för att säkerställa att alla regler upprätthålls (se NFER 2023).

På basis av utfallen på baslinjebedömningen, de nationella proven i årskurs 6 och examensskrivningarna i årskurs 11 beräknas sedan skolors förädlingsvärden från förskoleklass (tidigare årskurs 2) till årskurs 6 och från årskurs 6 upp till årskurs 11. Dessa förädlingsvärden offentliggörs årligen av det engelska utbildningsdepartementet och publiceras även i nationella tidningar. Sedan 2017 används förädlingsvärden som huvudsakligt mått på skolors prestationer, något som forskning visar fångar upp engelska skolors kausala effekter på inlärningen (Britton m.fl. 2023). Fokus på förädlingsvärden har också gjort att skolor inte enbart fokuserar på att stärka elever som ligger runt gränsen för godkänt (Burgess and Thomson 2023).

Över lag har systemet för kunskapsuppföljning- och information i England många fördelar och framstår som mer avancerat än det svenska. Även om man i flera fall har använt kunskapskontroller för olika syften har man också justerat dem för att de ska kunna fungera för dessa syften. Ett möjligt problem är dock att man, till skillnad från i Danmark, numera har ganska få kunskapskontroller under grundskoletiden, vilket minskar möjligheten till uppföljning över tid. Samtidigt har man skapat ett förhållandevis välutvecklat system för mätning av skolvalitet på basis av existerande kunskapskontroller.

5.3.USA

På grund av USA:s decentraliserade utbildningssystem finns det en mängd olika system för kunskapsuppföljning och information som varierar på delstatsnivå. På grundskolenivå finns inga gemensamma prov som alla elever skriver, men sedan den federala lagen "No Child Left Behind" (NCLB) infördes 2002 måste alla delstater (för att få federalt ekonomiskt stöd) anordna årliga standardiserade prov i läsning och matematik i årskurs 3-8 och i naturvetenskap en gång i årskurs 3-5 respektive 6-9 där alla elever deltar. Proven måste också skrivas på gymnasienivå. Syftet var framför allt att stärka kunskapsuppföljningen så att alla elever – speciellt de med utsatt bakgrund – kan nå godkända resultat genom att hålla skolor ansvariga för deras prestationer (Stullich m.fl. 2007).

Lagen hade stora effekter på delstaternas kunskapsuppföljningssystem. Före NCLB testade bara en fjärdedel av delstaterna eleverna i samma ämne minst två gånger mellan årskurs 2-3 och 8. I de flesta delstater var det i praktiken omöjligt att följa upp elevers prestationer systematiskt (Goldhaber och DeArmond 2023). Sedan tidigt 2000-tal har alltså de amerikanska utbildningssystemen kommit att inkludera mycket mer standardiserad kunskapsuppföljning på elevnivå än de gjorde tidigare. Exakt hur systemen fungerar varierar dock eftersom det är upp till delstaterna att själva skapa dem.

Det har funnits problem med NCLB, då de starka incitament som knyts till proven och resultaten via starkt ansvarsutkrävande – inklusive omstruktureringar av skolor som inte når målen – har lett till olika former av manipulation och snuttifiering av undervisningen. Trots detta tyder forskningen på att den har lett till högre kunskaper – även enligt kunskapskontroller som inte knyts till några incitament – men det är svårt att veta i vilken utsträckning detta handlar om kunskapsuppföljningarna i sig jämfört med själva ansvarsutkrävandet (se Goldhaber och DeArmond 2023; Heller-Sahlgren och Jordahl 2016).

Samtidigt har USA ett av de mest utvecklade systemen för urvalsbaserade kunskapsuppföljningar på nationell nivå i världen: "National Assessment of Educational Progress" (NAEP). Ända sedan tidigt 1970-tal har man utvärderat nio- och trettonåringar (samt sjuttonåringar) i läsning och matematik vart fjärde år. Under åren har systemet inkluderat tio ämnesområden: (1) dans, teater, musik och bild, (2) samhällskunskap, (3) ekonomi, (4) geografi, (5) matematik, (6) läsning, (7) naturvetenskap, (8) teknik, (9) amerikansk historia och (10) skrivning. Mest fokus har legat på läsning och matematik i vilka prov nu genomförs vartannat år i årskurs 4 och 8 sedan 1992 och vart fjärde år i de longitudinella åldersbaserade urvalen, som alltså går tillbaka till 1970-talet.²⁵ Likaså anordnas numera prov i samhällskunskap, amerikansk historia och naturvetenskap i årskurs 8 vart fjärde år. De andra ämnena testas mer sporadiskt (NCES 2024; SOU 2016:25).

Sedan NCLB infördes måste delstater också delta med ett representativt urval i NAEP i matematik och läsförståelse i både årskurs 4 och 8 för att få federal finansiering. Detta gör att man på rättvisande sätt kan jämföra hur olika delstater presterar, både jämfört med varandra och med sig själva över tid (NCES 2005; Goldhaber och DeArmond 2023). Dessutom deltar också drygt 20 större skoldistrikt på frivillig basis, vilket gör att man kan jämföra prestationerna mellan dessa.

Som noterades ovan ledde NCLB till att delstater i praktiken tvingades införa kontinuerliga kunskapsuppföljningar på basis av standardiserade prov i ett antal ämnen. Sedan dess måste också alla skoldistrikt publicera skolors absoluta resultat i årliga rapporter, jämförelser mellan skolor i distriktet och i USA generellt, samt information om vilka skolor som är i behov av förbättring. Fokus har framför allt legat på andelen som når kraven för godkänt, men man har under det senaste decenniet lagt större fokus på förädlingsvärden. Skolor har kunnat ansöka om att undantas från vissa krav i utbyte mot att lärare och rektorer utvärderas efter förädlingsvärden (Heller-Sahlgren och Jordahl 2016). Man har alltså börjat lägga större vikt vid förädlingsvärden på federal nivå än tidigare.

Vissa delstater har dock fokuserat på skolors renodlade förädlingsvärden under ganska lång tid. Ett exempel är North Carolina som sedan 1996 har hållit skolor ansvariga för årliga förändringar i elevers resultat i matematik och läsförståelse. Man har därefter beräknat förväntade förbättringar för varje elev efter deras tidigare resultat och skolor har klassats som effektiva eller ineffektiva beroende på om de lever upp till dessa förväntade förbättringar. Man har också givit ekonomiska incitament till skolorna att nå förväntningarna genom att ge bonusar till lärarna om skolorna klarar sig väl. Samtidigt har skolor som är ineffektiva offentliggjorts som sådana och ibland sätter myndigheterna in extra insatser i dessa skolor (Ladd och Lauen 2010). Sedan 2013 beräknar North Carolina också enskilda lärares förädlingsvärden med hjälp av ett system som utvecklats av företaget SAS Institute (se Lauen m.fl. 2013). Mer generellt mäter ett flertal delstater lärares effektivitet med hjälp av olika slags förädlingsvärden (se Walsh och Isenberg 2015). Lärares förädlingsvärden används alltså ofta som ett verktyg av myndigheter och skolor, men offentliggörs vanligtvis inte.

Över lag framstår systemet för kunskapsuppföljning- och information i USA som

Det finns också flera enskilda delstater där man beräknar skolors och lärares förädlingsvärden.

25: Eleverna i de åldersbaserade urvalen skriver prov med väldigt lika innehåll sedan 1970-talet, medan eleverna i de årskursbaserade urvalen skriver prov som uppdateras över tid för att reflektera läroplansförändringar (Goldhaber och DeArmond 2023). Även i de åldersbaserade urvalen har man dock gjort förändringar. Med hjälp av ankaruppgifter och "bryggstudier" har man säkerställt att provens svårighetsgrad på grundskolenivå inte varierar över tid och att eventuella andra förändringar inte påverkar möjligheterna till trendmätning, exempelvis efter att proven digitaliserades 2017 (se Jewsbury m.fl. 2020).

relativt sett avancerat. Kravet på kunskapskontroller i NCLB har gjort det möjligt att följa upp elever årligen i läsning och matematik, med det explicita syftet att stärka kunskapsutvecklingen och hålla skolor ansvariga. Kvalitetsinformation som skapas utifrån kunskapskontrollerna – och de incitament som knyts till dem – har över lag fokuserat för mycket på elevers absoluta resultat, men har blivit bättre med tiden. Trots evidens att systemet har lett till olika former av manipulation tycks nettoeffekten på inläringen ändå vara positiv. Det finns också flera enskilda delstater där man beräknar skolors och lärares förädlingsvärden och använder dessa i förbättringsarbetet. Dessutom har USA ett av de mest välutvecklade systemen för nationell och regional uppföljning i världen, vilket också inkluderar ämnen som normalt inte inkluderas i sådana undersökningar.

6. En vägkarta för bättre kunskapsuppföljning och information

I det här avsnittet presenteras en vägkarta för att skapa ett bättre system för kunskapsuppföljning och information i den svenska grundskolan. Tillsammans bör förslagen ses som skiss på ett system för kunskapsuppföljning på (1) individnivå, (1) skol- och huvudmannanivå samt (3) nationell nivå.

Som noterats i föregående avsnitt är det uppenbart att det i praktiken är omöjligt att skapa ett sådant system på basis av en typ av kunskapskontroll, eftersom kraven på hög validitet i alla tre hänseenden knappast går att kombinera i en enskild sådan. I stället krävs olika kunskapskontroller som utformas med olika syften i åtanke. Men det är också avgörande att utformningen av varje enskild typ av kunskapskontroll skapas med hela systemet i åtanke. Även om det är omöjligt att skapa ett perfekt system kan man på så sätt minimera problem med bristande validitet i olika hänseenden.

6.1. Inför årliga adaptiva prov i grundskolan

Jag föreslår att man skapar återkommande läroplansbaserade självvärtande prov i grundskolan i följande ämnen: svenska, matematik, SO-ämnena (ett prov) och NO-ämnena (ett prov). Digitala självvärtande prov minimerar den administrativa bördan för skolorna och innebär att resultaten är tillgängliga omedelbart efter att eleverna har skrivit provet. Proven bör skrivas årligen under tidig hösttermin från och med årskurs 1 till och med årskurs 9. Proven skulle därmed täcka alla teoretiska ämnena som eleverna får undervisning i under hela grundskolan.

Jag föreslår också att proven görs adaptiva.

Det huvudsakliga syftet med proven bör vara att mäta läroplansbaserade kunskaper och förändringar i dessa på elevnivå för att ge relevant information till elever, föräldrar och lärare. De sekundära syftena är att fungera som ett ankare för att säkerställa att kraven för olika resultat på de nationella proven är lika höga över tid (se avsnitt 6.2) och som mått för att skapa förädlingsvärden på skolnivå (se avsnitt 6.4).

Jag föreslår också att proven görs adaptiva.²⁶ Det finns en rad fördelar med

adaptiva prov, som diskuteras mer ingående i Appendixet.²⁷ Framför allt mäter adaptiva prov elevens resultat med större precision och även över ett större spann jämfört med vanliga linjära prov. Detta innebär att de är bättre på att fånga upp låg- och högpresterande elevers resultat jämfört med vanliga prov. Därför behöver elever generellt också svara på färre frågor i adaptiva än i traditionella prov för att uppnå samma validitet och reliabilitet: normala adaptiva prov är ofta ungefär 50 procent kortare utan att validiteten eller reliabiliteten påverkas (se Skolverket 2010b; Weiss och Kingsbury 1984).²⁸

Med hjälp av den senaste tekniken kan man även skapa så kallade multidimensionella adaptiva prov som med större säkerhet estimerar elevers utfall i flera delområden i samma prov, när elevers prestationer i dessa områden är korrelerade. Detta gör att man kan estimerar elevers kunskaper i olika delområden på kortaste möjliga tid och utan att skapa ett enskilt prov för varje separat delområde (se Braeken m.fl. 2020).

Det finns dock alltid en målkonflikt mellan total provtid och hur många delområden som kan testas med tillräcklig hög säkerhet, även om det är svårt att säga hur stor den är på förhand. Att låta elever skriva proven vid flera tillfällen gör att proven riskerar att bli alltför tidskrävande för skolorna. Rapporten föreslår därför att man enbart har ett provtillfälle i varje ämne med normal provlängd och inom ramen för denna begränsning använder multidimensionell teknik för att estimerar förmågor i olika delområden. Idén med proven är att de huvudsakligen ska användas för att ge information om elevers övergripande kunskapsutveckling på ett objektivt sätt.

Proven skulle för tillfället sannolikt huvudsakligen, givet nuvarande tillgänglig teknik, vara begränsade till flervalsfrågor och frågor där elever anger korta svar – eftersom frågorna måste rättas i realtid allteftersom eleverna svarar på dem – även om utvecklingen inom artificiell intelligens vittnar om att man sannolikt snart kommer att kunna inkludera frågor där eleverna får svara mer fritt.²⁹ Givet att bedömningen av elevlösningar och muntliga delar på de nationella proven, där eleverna i högre utsträckning får svara på frågor som kräver friare svar, lider av låg samstämmighet mellan bedömare är fördelarna med att elever får svara fritt sannolikt mer skenbar än verklig eftersom bedömningen lider av subjektivitet (se t.ex. Skolinspektionen 2021).³⁰ Forskning visar dessutom att man oftast fångar upp samma slags kunskaper med flervalsfrågor och öppna frågor, vilket ytterligare vittnar om att de påstådda fördelar som finns med de sistnämnda är mer skenbar än verklig (se Hift 2014).

26: Eftersom proven skrivs i början av hösten bör de adaptiva proven ställas in så att eleverna börjar provet på en genomsnittlig svårighetsgrad för elever i föregående årskurs (se Appendix).

27: Det finns naturligtvis även nackdelar, såsom svårigheter att förstå proven och hur resultaten beräknas, vilka alltså gjorde att Danmark valde att ta bort den adaptiva funktionen. Jag bedömer dock att fördelarna är många fler än nackdelarna givet det tänkta huvudsakliga syftet med proven: att mäta kunskaper och förändringar i detta hänseende. Man bör också notera att andra länder under de senaste åren tvärtom har gjort sina provsystem adaptiva på grund av fördelarna. Exempel inkluderar Australien och USA där NAPLAN-proven respektive SAT-proven (det amerikanska högskoleprovet) har gjorts adaptiva i och med att man digitaliserade dem (se Meltzer 2023; NAP 2024).

28: Eftersom alla prov är unika kan man säga att de egentligen inte är prov i traditionell mening, utan snarare en stor uppgiftsbank från vilken en algoritm väljer ut specifika uppgifter som elever får svara på under ett visst bedömningstillfälle. Varje elevs unika prov skapas alltså i praktiken automatiskt av algoritmen i realtid, på basis av hur eleverna svarar, även om man måste ha bestämt vilka gränsvärden som ska gälla för olika provbetyg genom "standard setting", enligt diskussionen nedan.

29: Redan nu skulle man sannolikt kunna utveckla automaträttade essäfrågor för att bedöma elevers skrivförmåga. Ett sådant system har redan skapats i Australien för att automaträtta essäfrågor i engelska, även om det ännu inte har införts av politiska skäl (Lazendic m.fl. 2018; Robinson 2018). Över lag finner dock forskning fortfarande utmaningar med sådana system (Ramesh och Sanampudi 2022). Analyser har också intressant nog funnit att bedömningen av essäfrågor med ChatGPT tycks vara ganska mycket sämre än människors bedömningar (se t.ex. Wheadon 2023). Det är också värt att undersöka möjligheterna att använda mindre konventionella tester. Ett exempel är företaget Lexplores metod att mäta läsförmåga med ögonrörelser och artificiell intelligens.

30: Att frågorna i de adaptiva proven huvudsakligen kommer att bestå av flervalsfrågor och frågor där elever ger korta svar betyder inte att alla svar enbart måste bedömas som rätt eller fel. Uppgifter i adaptiva prov kan utformas på ett sätt där man får poäng för ett svar som indikerar viss förståelse (så kallad "partial credit").

Det finns dock alltid en målkonflikt mellan total provtid och hur många delområden som kan testas med tillräcklig hög säkerhet.

En viktig fråga är om rättningen av proven ska vara normrelaterad (relativ) eller målrelaterad. Med relativa prov skulle de adaptiva proven fånga upp hur elever presterar i den nationella fördelningen och hur deras position i denna fördelning förändras över åren. Det finns fördelar med sådana prov, då formatet kan variera från år till år utan att man behöver säkerställa att proven är exakt lika svåra. Nackdelen är att man då inte kan fånga upp i vilken utsträckning elever når målsättningarna i läroplanen. I det här fallet är fördelen att ändra provformatet över tid också inte speciellt relevant, givet de slags frågor som för tillfället kan inkluderas.

Eftersom proven är ämnade att vara en kontroll av elevernas kunskaper i relation till läro- och kursplaner är det sannolikt bättre att göra dem målrelaterade och rätta dem enligt en kontinuerlig skala som motsvarar den för nationella prov (just nu 0–20). Till skillnad från i de nationella proven skulle dock hela skalan kunna användas.³¹ De olika gränsvärdena bestäms med så kallad ”standard setting” (se Sireci m.fl. 2000).

Eftersom adaptiva prov bygger på att elever ska kunna svara på frågor med olika svårighetsgrad måste man estimeras svårighetsgraden för varje enskild fråga i provkonstruktionen. Detta görs med hjälp av pilotprov. När nya frågor läggs in måste deras svårighetsgrad också estimeras på samma sätt.³² På detta sätt säkerställer man också att proven är lika svåra över tid (se t.ex. He 2010).

Inga krav bör finnas för att proven ska användas i betygssättningen. Detta för att minska incitamenten att undervisningen enbart börjar riktas mot proven. Däremot bör det inte finnas något hinder för lärare att inkludera utfallen som grund för betygssättningen, på frivillig basis, i den mån de själva ansvarar för den.³³ Provresultaten är av relevans för att förstå i vilken utsträckning elever når upp till olika kunskapskrav och om lärare vill använda den informationen i betygssättningen är det positivt. Att proven skulle kunna användas för det syftet gör också att elever har en anledning att bry sig om dem.³⁴

6.2. Låt de (externt rättade) nationella provens utformning variera över tid och använd de adaptiva proven som nivåankare

Som noterats i tidigare avsnitt fyller de nationella proven i årskurs 6 och 9 idag inte sitt huvudsakliga syfte: att fungera som ett instrument för likvärdig betygssättning. Just nu pågår en offentlig utredning som ska lämna förslag på hur man kan stärka likvärdigheten genom att stärka kopplingen mellan resultaten på de nationella proven och betygen i slutet på grundskolan (Dir. 2023:95). På något sätt kommer alltså betygssättningen att kopplas till de nationella proven (som även kan göras om till examensprov), exempelvis genom att man knyter

31: På de nationella proven skapas provbetygspoängen utifrån vilket betyg eleverna uppnår: F ger 0 poäng, E ger 10 poäng, D ger 12,5 poäng, C ger 15 poäng, B ger 17,5 poäng och A ger 20 poäng.

32: Nya frågor kan dock, om man så vill, blandas in i uppgiftsbanken från vilken de adaptiva proven skapas i realtid och därmed kalibreras samtidigt som eleverna skriver dem. Elevernas provresultat beräknas dock då endast på basis av uppgifter som är färdigställda som en del av den kalibrerade uppgiftsbanken. Detta liknar modellen som används för högskoleprovet, där en del av provet är en utprövningsdel som inkluderar ankarfrågor utan att kandidaterna vet om att denna del inte påverkar deras resultat (se Wiberg 2019).

33: Eftersom betygssättningen med stor sannolikhet, på något sätt, kommer att knytas till resultaten på nationella prov innebär detta att de adaptiva proven framför allt skulle kunna fungera som ett hjälpmedel för lärarna för att avgöra hur betygen ska fördelas inom skolan. Detta skulle dock vara irrelevant för slutbetygen i årskurs 9 om man går över till en renodlad examensprovmodell.

34: Att proven påverkar skolornas förädlingsvärden gör att lärare och skolor mer generellt kommer att ha starka incitament att få eleverna att bry sig om proven.

skolors genomsnittliga meritvärden till deras genomsnittliga resultat på de (externt rättade) nationella proven (se Skolverket 2020). Vilken modell man än väljer kommer de nationella proven med stor sannolikhet att få en ännu viktigare roll för betygssättningen än idag.

Detta innebär också att incitamenten som knyts till dessa prov med stor sannolikhet kommer att öka markant.³⁵ Detta gör att man i sin tur får anta att fokus på proven kommer att öka ännu mer, vilket kan komma att leda till förändrad undervisning och försök att på olika sätt manipulera resultaten. Extern rättning är ett grundkrav men som situationen i England visar är risken stor att det kommer uppstå betygsinflation även med sådan rättning, på grund av kombinationen av starka incitament och svårigheterna att säkerställa att proven håller samma nivå över tid. Om detta sker kommer det i sin tur att spilla över på betygen som proven knyts till. Det kommer därför krävas åtgärder som minskar incitamenten att enbart fokusera på det som mäts på proven och som säkerställer att kraven för olika resultat på de nationella proven är lika höga över tid.³⁶

Rapporten rekommenderar därför att man skapar ett antal olika prov vars format till viss del skiljer sig åt men som skapas på ett sätt som gör att de är lika svåra. Provformatet bör alltså inte vara exakt samma varje år. I stället kan proven skapas enligt ett antal olika specifikationer. Exakt vilken specifikation som ges under vilket år kan sedan slumpas ut strax innan proven skrivs – formatet roterar således mellan åren. På detta sätt blir det svårt för aktörerna i skolsystemet att enbart träna på det format som de utvärderas efter, samtidigt som provformatet ändå inte är helt främmande för dem.

Naturligtvis måste man säkerställa att dessa prov är lika svåra, både mellan specifikationer och över tid.³⁷ Och i den mån de inte är lika svåra måste man ta hänsyn till det i bedömningen. Här rekommenderar jag att man hämtar inspiration från England, där man har stor erfarenhet av kalibreringar av externt rättade prov för att säkerställa att kraven inte skiljer sig över tid, trots förändringar i proven.

Sådan kalibrering kan göras på olika sätt, exempelvis med ankar- eller referensprov. Med sådana justeras kraven för olika resultat på de nationella proven på basis av hur ett urval elever presterar på ankar- eller referensprov, som innehåller uppgifter som inte förändras över tid, och på basis av relationen mellan prestationerna och resultaten på de nationella proven. Jag föreslår att man använder de adaptiva proven för detta syfte, på ett liknande sätt som man har gjort i England. De adaptiva proven kommer att skrivas i ett flertal ämnen årligen och kalibreras för att vara lika svåra över tid, vilket gör att man har ett stort data-material från varje årskull med vilket man kan säkerställa att kraven för olika resultat på de nationella proven inte varierar från år till år.

Detta betyder att relationen mellan resultaten på de adaptiva proven och de nationella proven inte kan ändras över tid på nationell nivå. Om kunskaperna totalt sett ökar/sjunker enligt de adaptiva proven ökar/sjunker också resultat-

35: Exakt vilken modell man väljer kommer att skapa olika incitament på elev- lärar- och skolnivå. Om man exempelvis gör om de nationella proven till renodlade examensprov kommer incitamenten att vara starka på individ-, lärar- och skolnivå. Om man normerar betygen efter skolors genomsnittliga resultat på nationella prov kommer incitamenten vara lika starka på skolnivå (och sannolikt på lärarnivå), men inte lika starka på elevnivå.

36: Det kommer också krävas ett system som säkerställer att alla elever gör proven och som tar bort möjligheterna att fuskas, exempelvis med extern provövervakning (se Heller-Sahlgren 2019). Detta är dock något som den pågående utredningen kommer att ta hänsyn till.

37: Ett annat enklare alternativ vore att göra resultaten relativa, vilket gör att det inte spelar någon roll exakt hur svårt ett givet prov är över tid. Men detta skulle även göra betygssystemet helt relativt, vilket det sannolikt inte finns någon aptit för.

en på de nationella proven totalt sett. Exakt vilka adaptiva provresultat man kalibrerar mot bör utredas vidare, men det vore rimligt om det är ett snitt över ett antal år.³⁸

Utredningen som nu pågår analyserar de nationella provens roll för betygssättningen. Det är dock oklart om förslagen enbart kommer beröra betygen i årskurs 9. Rapporten föreslår hursomhelst att ovanstående justeringar även appliceras på de nationella proven i årskurs 3 och 6. Dessa prov bör således rättas externt och formatet förändras över tid. I årskurs 3 innebär det också att man gör om proven till att inte enbart fokusera på minimikrav.

6.3. Inför ett system för trendmätning på nationell nivå

Jag rekommenderar att Sverige följer USA och skapar ett urvalsbaserat provsystem för att systematiskt studera hur kunskaperna förändras på nationell nivå. Liksom i USA kan prov i läsförståelse och matematik skrivas vartannat år, men om möjligt vore det även bra att genomföra prov vartannat år i NO- och SO-ämnena. Proven kan förslagsvis skrivas i årskurs 3, 6 och 9. Eftersom syftet med proven enbart är att mäta kunskapsutvecklingen på nationell nivå kan de utformas så att de täcker alla delar av ämnet som man anser vara viktiga, på ett liknande sätt som i internationella undersökningar (se t.ex. SOU 2014:12).

Då proven ska kunna användas för att följa upp den nationella kunskapsutvecklingen under lång tid framöver måste de utformas på ett liknande sätt som de internationella undersökningarna, där man ofta introducerar nya frågetyper och justerar provformatet men fortfarande kan jämföra utfallen över tid (se t.ex. OECD 2023).

Urvalsproven bör inte ha några explicita incitament alls knutna till sig på lärar- eller skolvivå, för att undvika att skolor och lärare börjar fokusera på att träna elever inför proven. Man bör däremot överväga att ge incitament till åtminstone ett urval av eleverna att göra så bra som möjligt ifrån sig. Sådana incitament bör i så fall introduceras strax innan provtillfället, för att säkerställa att de endast påverkar elevernas ansträngningar på provet. Forskning har visat att sådana incitament kan ha stora effekter på prov som annars inte spelar någon roll för eleverna (se Gneezy m.fl. 2019).

Nationella urvalsbaserade stickprov skulle vara ett viktigt komplement till de heltäckande kunskapsutvärderingar som utgörs av de adaptiva och nationella proven. Med hjälp av stickproven kan man också identifiera eventuella avvikelser mellan de olika kunskapsbedömningarna, vilka sedan skulle kunna utforskas närmare.³⁹

Jag rekommenderar att Sverige följer USA och skapar ett urvalsbaserat provsystem för att systematiskt studera hur kunskaperna förändras på nationell nivå.

³⁸: Man kan naturligtvis också, precis som England, även kombinera modellen för kalibrering med separata urvalsbaserade referensprov i ämnen som inte inkluderas i det adaptiva provsystemet.

³⁹: Man kan exempelvis analysera rimligheten i den kalibrering som sker av betygssättningen i de nationella proven med hjälp av de adaptiva proven. Detta på basis av idén att resultatutvecklingen på de adaptiva proven och stickproven totalt sett inte bör inte vara extremt olika.

6.4. Skapa relevant information på basis av kunskapskontrollerna

De individuella provresultaten på de adaptiva proven bör göras tillgängliga för rektorer, lärare, elever, föräldrar och myndigheter. På detta sätt får alla relevanta aktörer information om elevernas resultat och hur de förändras över tid. Däremot bör inte skolors genomsnittliga resultat på de adaptiva proven publiceras. Myndigheter bör dock använda informationen för att skapa förädlingsvärden på skolnivå, som sedan kan publiceras. I Appendixet redovisas en metod med vilka dessa värden kan skapas på översiktlig nivå.

Förädlingsvärdena som skapas utifrån de adaptiva proven bör beräknas som det oviktade genomsnittet av skolors årliga förädlingsvärde, mätt över alla enskilda år. Detta skulle minska incitamenten för ett visst slags manipulation: att man lägger sig på en låg nivå initialt för att göra det lättare att höja förädlingsvärdet senare (se t.ex. Macartney 2016).⁴⁰

Att man skapar förädlingsvärden på basis av de adaptiva proven skapar visserligen incitament att styra undervisningen mot dem, men detta är inte ett speciellt stort problem i det här fallet. I alla årskurser där de nationella proven skrivs – även i årskurs 3 – kommer nämligen de provens roterande format att fungera som ett incitament för skolor att inte börja snuttifiera undervisningen för att öka resultaten på de adaptiva proven. Detta helt enkelt eftersom sådan snuttifiering i sin tur riskerar att försämra resultaten på de nationella proven.

Utöver detta kommer de nationella proven också ha speciellt starka incitament knutna till sig i årskurs 6 och 9. Detta eftersom de kommer att avgöra betygssättningen. Att fokusera all kraft och energi på att träna elever för adaptiva prov som inte avgör skolornas genomsnittliga betyg – vars format kommer att skilja sig från de prov som kommer att vara avgörande för betygssättningen – vore inte speciellt strategiskt.

Provresultaten på de nationella proven bör också göras tillgängliga på individ- och skolnivå på samma sätt som de adaptiva proven. Detta innebär att man även kan skapa ytterligare ett förädlingsvärde på basis av resultatet från de första adaptiva proven som eleverna skriver till nästa nationella prov som skrivs. Fördelen är att man då skapar förädlingsvärdet på basis av ett utfallsmått som alla aktörer, elever, skolor och lärare, har starka incitament att prestera bra på – samtidigt som det roterande provformatet alltså minskar möjligheterna att snuttifiera undervisningen efter utfallsmåttet. Detta speciellt då provresultaten kommer att justeras på basis av resultaten på de adaptiva proven för att säkerställa en jämn standard över tid.

Med rätt systemdesign finns det således inga större skäl för skolor och lärare att lägga allt fokus på vare sig de adaptiva proven eller de nationella proven, trots att man använder dem för att skapa och publicera förädlingsvärden. De olika proven fungerar helt enkelt som en slags spärr för att skolor inte ska ha alltför starka incitament att snuttifiera undervisningen i en specifik riktning.

De olika proven fungerar helt enkelt som en slags spärr för att skolor inte ska ha alltför starka incitament att snuttifiera undervisningen i en specifik riktning.

⁴⁰: Incitament att manipulera ingångsresultaten minskar eftersom man bedöms på basis av ingångsvärden som på samma gång blir utfallsvärden året efteråt. Om ett högstadium presterar dåligt på provet i både årskurs 7 och 8 men sedan jättebra på provet i årskurs 9 kommer genomsnittet av förädlingsvärdena vara lika bra som om man höjde resultaten jämnt under alla år. Och om skolan enbart presterar sämre på provet i årskurs 7 och sedan presterar jättebra på provet i årskurs 8 påverkar det i sin tur möjligheterna att höja resultaten i årskurs 9.

Eftersom man inte vill skapa för många mått som ämnar fånga samma sak skulle man kunna skapa ett totalt förädlingsvärde på basis av genomsnittet av värdet som skapas utifrån enbart de adaptiva proven och det som skapas utifrån adaptiva prov samt nationella prov.

Tekniskt sett är det inte strikt nödvändigt att resultaten på de nationella proven publiceras på skolnivå, givet att betygen på ett eller annat sätt kommer att knytas till dessa provresultat. Men det är rimligt att göra det för att säkerställa transparens. Dock bör alltid förädlingsvärdet vara det huvudsakliga måttet på skolkvalitet.

Vad gäller resultaten på de nationella trendmätningarna bör dessa enbart publiceras och analyseras (förutom i oidentifierad form) på nationell nivå. Detta för att säkerställa att inga incitament alls knyts till kunskapsmätningarna på skol- eller lärarnivå. Som mest kan man i efterhand visa skolors genomsnittliga resultat för rektorerna, om de så vill, med garanti för att resultaten hemlighålls.

6.5. Inrätta ett bedömningscentrum med ansvar för all kunskapsuppföljning

Ett oberoende bedömningscentrum som är ansvarigt för all kunskapsuppföljning och analys bör inrättas. Ett sådant centrum kan med fördel förläggas vid ett universitet och bör knyta till sig stark psykometrisk och nationalekonomisk kompetens.

Detta betyder inte att centrumet bör skapa alla prov och förädlingsvärdena, men att de ska vara ansvariga för det övergripande systemet som har skissats ovan. På detta sätt frikopplas bedömningen från Skolverket, även om samarbeten säkerligen kommer att skapas.

Ovanstående vägkarta är enbart en skiss. Innan man bestämmer sig för exakt hur proven och måtten bör utformas rekommenderar rapporten att centrumet tillsätter en expertgrupp av forskare med uppdrag att analysera olika slags prov och informationsmått.

7. Slutsatser

Den här rapporten har diskuterat bristerna i systemet för kunskapsuppföljning och information i den svenska grundskolan och föreslagit reformer för att förbättra situationen. Den har dragit lärdomar från forskningen kring effekterna av kunskaps- och kvalitetsuppföljning och även diskuterat hur andra länder har utformat relevanta system.

Forskning visar att kunskapsuppföljningar och information som skapas från dem ofta höjer elevers kunskaper. Information om skolkvalitet – och ansvarsutkrävande på basis av denna information – har också ofta positiva effekter.

Rapporten har dragit lärdomar från Danmark, England och USA för att förbättra det svenska systemet för kunskapsuppföljning och information. Till att börja med bör man införa årliga läroplansbaserade självvärtande och adaptiva prov i grundskolan i svenska, matematik, SO-ämnena och NO-ämnena. Dessutom bör de (snart externt rättade) nationella provens utformning variera över tid, med de adaptiva proven som nivåankare för att säkerställa likvärdiga krav för olika betyg över tid.

Likaså bör man skapa ett urvalsbaserat provsystem för att systematiskt studera hur kunskaperna förändras på nationell nivå. Prov i läsförståelse och matematik samt SO- och NO-ämnena bör skrivas vartannat år bland elever i årskurs 3, 6 och 9.

Individuella resultat på de adaptiva och nationella proven bör göras tillgängliga för rektorer, lärare, elever, föräldrar och myndigheter. Myndigheter kan sedan använda informationen för att skapa förädlingsvärden på skolnivå. Genom samspelet mellan de adaptiva respektive nationella proven begränsas skolornas incitament och möjligheter att enbart styra undervisningen i endera riktningen, trots att man använder proven för att skapa information om skolors kvalitet. Resultaten på de nationella trendmätningarna bör dock enbart publiceras och analyseras (förutom i avidentifierad form) på nationell nivå.

Till sist föreslog rapporten att man inrättar ett oberoende bedömningscentrum som är ansvarigt för all kunskapsuppföljning och analys. Ett sådant centrum kan med fördel förläggas vid ett universitet och bör knyta till sig stark psykometrisk och nationalekonomisk kompetens.

Väggkartan som föreslås i rapporten är naturligtvis enbart en grov skiss. Det bedömningscentrum som inrättas bör därför tillsätta en expertgrupp med uppdrag att analysera olika slags prov och informationsmått för att avgöra den exakta vägen framåt.

**Väggkartan
som föreslås i
rapporten är
naturligtvis
enbart en
grov skiss.**

Appendix

Adaptiva prov

Att mäta elevers progression kan underlättas om man skapar proven efter den psykometriska forskningen kring så kallade adaptiva prov (se t.ex. Skolverket 2010b; Weiss och Kingsbury 1984). I stället för att alla elever får samma frågor – som i vanliga prov (som också kallas ”linjära” prov) används en algoritm för att under provets gång välja vilka frågor – från en i förväg sammanställd uppgiftsbank – som varje enskild elev ska besvara. Eftersom alla prov är unika kan man säga att de egentligen inte är prov i traditionell mening, utan snarare en stor uppgiftsbank från vilken algoritmen väljer ut specifika uppgifter som elever får svara på under ett visst bedömningstillfälle. Varje elevs unika prov skapas alltså i praktiken automatiskt av algoritmen på basis av hur eleverna svarar i realtid, även om man måste ha bestämt vilka gränsvärden som ska gälla för de olika åren genom ”standard setting” (se Sireci m.fl. 2000).

Baserat på elevens svar på en fråga beräknas elevens förmåga (som kan motsvara ett specifikt betygssteg eller annan form av bedömningssteg), som sedan används för att bestämma vilken efterföljande fråga som denne ska besvara. Elever som klarar en specifik uppgift får en svårare efterföljande fråga, medan elever som inte klarar uppgiften får en lättare efterföljande fråga.

På basis av svaren uppdateras sedan estimatet av elevens förmåga under provets gång och nya frågor presenteras fram tills dess att ett så kallat slutkriterium har uppnåtts. Detta slutkriterium kan baseras på olika villkor. Eftersom målet är att estimeras elevens förmåga avslutas ofta provet när felmarginalen för den uppmätta förmågan faller under ett förutbestämt värde.

Eftersom adaptiva prov bygger på att elever ska kunna svara på frågor med olika svårighetsgrad måste man estimeras svårighetsgraden för varje enskild fråga i provkonstruktionen. Detta görs med hjälp av pilotprov. När nya frågor läggs in måste deras svårighetsgrad också estimeras på samma sätt. På detta sätt säkerställer man också att proven är lika svåra över tid (se t.ex. He 2010). Nya frågor kan dock, om man så vill, blandas in i uppgiftsbanken från vilken de adaptiva proven skapas i realtid och därmed kalibreras samtidigt som eleverna skriver dem. Elevernas provresultat beräknas dock då endast på basis av uppgifter som är färdigställda som en del av den kalibrerade uppgiftsbanken. Detta liknar modellen som används för högskoleprovet, där en del av provet är en utprovingsdel som inkluderar ankarfrågor utan att kandidaterna vet om att denna del inte påverkar deras resultat (se Wiberg 2019).

Jämfört med linjära prov mäter adaptiva prov elevens resultat med större precision och även över ett större spann. Dessa fördelar kommer av att elever endast svarar på frågor som är lämpliga givet deras förmåga. I linjära prov kan man generellt inte säkerställa att det finns tillräckligt många frågor för att mäta alla elevers nivåer med ett enda prov. Därför behövs det också generellt färre frågor i adaptiva än i traditionella prov för att uppnå samma validitet och reliabilitet: digitala adaptiva prov kan vara cirka 50 procent kortare.

Med hjälp av den senaste tekniken kan man även skapa så kallade multidimensionella adaptiva prov som med större säkerhet estimerar elevers utfall i flera delområden i samma ämne, när elevers prestationer i dessa delområden är på kortaste möjliga tid och utan att skapa ett enskilt prov för varje separat delområde (se Braeken m.fl. 2020).

En annan fördel är att de adaptiva proven möjliggör mätningar av hur elevers kunskaper förändras över tid eftersom detta kan utvärderas med samma slags prov, men utan att eleverna behöver svara på samma frågor som de gjorde vid första provtillfället. Dessutom minskar risken för fusk och effekterna av läckta provfrågor minimeras, eftersom elever omöjligen kan veta vilka frågor just de kommer att få på proven. Detta gör också att elever kan skriva proven vid ett senare tillfälle om de råkar vara frånvarande av någon anledning.

Eftersom systemet kräver att elevers svar rättas i realtid krävs ingen rättning från lärare. Då proven rättas automatiskt när de skrivs kan man också säkerställa nästan omedelbar återkoppling till både lärare och elever, vilket är avgörande för att man ska kunna reagera på den information som proven ger.

Metod för att beräkna förädlingsvärden

En enkel metod för att beräkna förädlingsvärden i den svenska kontexten finns beskrivna i exempelvis Holmlund m.fl. (2019), med följande generella regressionsmodell:

$$P_{is}^c = \alpha V_{is}^{c-n} + \beta X_i + \mu_s + \varepsilon_{is}$$

där P_{is}^c är valt kunskapsutfall P för elev i vid tidpunkt c på skola s , medan V_{is}^{c-n} är valt kunskapsutfall V för samma elev i vid tidpunkt $c-n$. Vektorn X_i inkluderar de bakgrundsvariabler man vill hålla konstant i modellen.⁴¹ Ekvationen innehåller även en skolnivåeffekt, μ_s , och en felterm, ε_{is} .⁴²

Förädlingsvärdet för skola j är enligt denna modell:

$$FV_j = \frac{1}{n_j} \sum_{i \text{ om } s=j} (P_{is}^c - (aV_i^{c-n} + bX_i)),$$

där n_j är antalet elever på skola j . På basis av detta kan man också skapa ett konfidensintervall för att säkerställa att skolor skiljer sig med en given säkerhetsgrad. Detta kan vara 95 procent – vilket är den normala gränsen för att en skillnad ska sägas vara statistiskt säkerställd – men i flera fall kan det vara rimligt att använda en lägre säkerhetsgrad. Så länge skolor skiljer sig åt med mer än 50 procents sannolikhet är det exempelvis bättre att använda måttet vid beslut än att inte använda det (Heller-Sahlgren och Jordahl 2016).

Beräkningarna av skolkvalitet är per definition relativa, vilket betyder att skolornas förädlingsvärden ska tolkas i relation till varandra. En skola vars elever presterar väl enligt kunskapskontrollen som används som baslinjeprov och som försämrar sina resultat till den andra kunskapskontrollen kan ändå ha relativt högt förädlingsvärde – om elever på andra skolor har försämrat sina resultat ännu mer.⁴³

41: Engelsk forskning tyder på att man inte behöver ha med bakgrundsfaktorer över huvud taget när man har tillgång till elevernas kunskapsresultat vid nollåget. Detta trots att det är fem år mellan det första och andra mätillfället och att det första mätillfället inte berör alla ämnen som inkluderas i utfallsmåttet (se Britton m.fl. 2023).

42: Ett annat alternativ är att använda en icke-parametrisk approach, likt den som används i England. I den versionen delas elever in i 34 lika stora grupper på basis av deras tidigare resultat. Förädlingsvärdet på elevnivå är sedan deras eget utfall minus det genomsnittliga utfallet i den grupp som eleven tillhör. Forskning tyder på att det inte går att statistiskt skilja den metoden från den som föreslås ovan (Britton m.fl. 2023), men det är värt att undersöka om så är fallet i den svenska kontexten.

43: Det går att justera modellen för att den ska fånga upp absoluta förändringar. För att det ska vara möjligt måste man enbart inkludera tidigare årskullar i modellen, med vilken man sedan förutspår resultaten bland elever i en senare årskull. Fördelen är att alla skolor tekniskt sett då kan slå sina modellberäknade värden.

Referenser

Allerup, Peter och Christian Christrup Kjeldsen. 2017. "Standard setting in Sweden: Challenges Through Computer-Based Adaptive Testing." I Sigrid Blömeke och Jan-Eric Gustafsson (red.). *Standard Setting in Education: The Nordic Countries in an International Perspective*. Springer.

Andersen, Simon Calmar och Helena Skyt Nielsen. 2020. "Learning from Performance Information" *Journal of Public Administration Research and Theory* 30(3): 415–431.

Azmat, Ghazala och Nagore Ireberri. 2010. "The importance of relative performance feedback information: Evidence from a natural experiment using high school students" *Journal of Public Economics* 94(7–8): 435–452.

Beach, Phil. 2015. "GCSE marking and grading." Blogg, Ofqual, <https://ofqual.blog.gov.uk/2015/08/05/gcse-marking-and-grading/>.

Benton, Tom. 2016. "Comparable outcomes: Scourge or scapegoat?" Rapport, Cambridge Assessment.

Benton, Tom och Tom Sutch. 2014. "Analysis of the use of Key Stage 2 data in GCSE predictions." Rapport, Cambridge Assessment.

Bergbauer, Annika B., Erica Hanushek och Ludger Woessmann. 2024. "Testing" *Journal of Human Resources* 59(2): 349–388.

Bergman, Peter. 2019. "How Behavioral Science Can Empower Parents to Improve Children's Educational Outcomes." *Behavioral Science & Policy* 5(1): 1–91.

Bergman, Peter. 2021. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment" *Journal of Political Economy* 129(1): 286–322.

Betts, Julian R., Youjin Hahn och Andrew C. Zau. 2027. "Learning from Performance Information" *Journal of Urban Economics* 100: 54–64.

Beuchert, Louise, Tine Louise Mundbjerg Eriksen och Morten Visby Krægpøth. 2020. "The impact of standardized test feedback in math: Exploiting a natural experiment in 3rd grade" *Economics of Education Review* 77: 102017.

Beuchert, Louise V. och Anne B. Nandrup. 2020. "The Danish National Tests at a Glance" *Nationaløkonomisk tidsskrift* 2018(1): 1–37.

Bramley, Tom och Carmen Vidal Rodeiro. 2014. "Using statistical equating for standard maintaining in GCSEs and A levels." Rapport, Cambridge Assessment.

Breakspear, Simon. 2012. "The Policy Impact of PISA: An Exploration of the

Normative Effects of International Benchmarking in School System Performance". OECD Education Working Papers No. 71.

Braeken, Johan och Muirne C. S. Paap. 2020. "Making Fixed-Precision Between-Item Multidimensional Computerized Adaptive Tests Even Shorter by Reducing the Asymmetry Between Selection and Stopping Rules". *Applied Psychological Measurement* 44(7-8): 531-547.

Britton, Jack, Damon Clark och Ines Lee. 2023. "Exploiting discontinuities in secondary school attendance to evaluate value added". IFS Working Paper 23/24.

Burgess, Simon och Dave Thomson. 2023. "Beyond the threshold: The implications for pupil achievement of reforming school performance metrics." *Economics of Education Review*, 97: 102477.

Burgess, Simon, Deborah Wilson och Jack Worth. 2013. "A natural experiment in school accountability: The impact of school performance information on pupil progress and sorting." *Journal of Public Economics*, 106: 57-67.

Claeson, Peder. 1987. "Använd standardproven i matematik...och diskutera efteråt!". *Nämnamn* 2-3, 11.

Cobb-Clark, Deborah A., Tiffany Ho och Nicolás Salamanca. 2021. "Parental Responses to Children's Achievement Test Results". IZA Discussion Paper No. 14663.

Coe, Robert. 2013. "Improving education: A triumph of hope over experience." Rapport, Centre for Evaluation & Monitoring, Durham University.

College Board. 2023. "What is Digital SAT Adaptive Testing?". Hemsida, <https://blog.collegeboard.org/what-digital-sat-adaptive-testing>.

Davoli, Maddalena och Horst Entorf. 2018. "The PISA Shock, Socioeconomic Inequality, and School Reforms in Germany". IZA Policy Paper No. 140.

Dir. 2023:95. "Likvärdiga betyg och meritvärden." Kommittédirektiv, <https://www.regeringen.se/contentassets/c307185c6f8a4726a091d29cad418fff/likvardiga-betyg-och-meritvarden-dir.-202395.pdf>.

Dizon-Ross, Rebecca. 2019. "Parents' Beliefs about Their Children's Academic Ability: Implications for Educational Investments." *American Economic Review* 109(8): 2728-2765.

Edin, Per-Anders, Peter Fredriksson, Martin Nybom och Björn Öckert. 2022. "The Rising Return to Noncognitive Skill." *American Economic Journal: Applied Economics* 14(2): 78-100.

Eurydice. 2024. "Assessment in single-structure education", Hemsida, <https://eurydice.eacea.ec.europa.eu/national-education-systems/denmark/assessment-single-structure-education>.

Fischer, Mira och Valentin Wagner. 2023. "Do timing and reference frame of feedback influence high-stakes educational outcomes?" *Economics of Education Review* 94: 102379.

Folkeskoleforligskredsen. 2021. "Aftale om det fremtidige evaluerings- og bedømmelsessystem i folkeskolen". Rapport, <https://www.uvm.dk/-/media/filer/uvm/aktuelt/pdf21/okt/211029-aftale-om-det-fremtidige-evaluerings-og-bedoemelsessystem-i-folkeskolen.pdf>.

Gneezy, Uri, John A. List, Jeffrey A. Livingston, Xiangdon Qin, Sally Saddoff och Yang Xu. 2023. "Measuring Success in Education: The Role of Effort on the Test Itself?" *American Economic Review: Insights* 1(3): 291–308.

Goldhaber, Dan och Michael DeArmond. 2023. "What Does Empirical Research Say About Federal Policy From NCLB to ESSA?". Rapport, U.S. Chamber of Commerce Foundation, Washington DC.

He, Qingping. 2010. "Maintaining standards in on-demand testing using item response theory". Rapport, Ofqual, London.

Heller-Sahlgren, Gabriel. 2019. "Förädlingsvärdets värde – att mäta skolors och lärares effektivitet". Rapport, Svenskt Näringsliv.

Heller-Sahlgren, Gabriel. 2022. "Så räddades skolan – men invandringen kräver mer". Kvartal, <https://kvartal.se/artiklar/den-svenska-skolan-ar-battre-an-sitt-rykte/>.

Heller-Sahlgren, Gabriel. kommande. "Tidiga och senare utfall i grundskolan – att skapa empiriska kunskapsprofiler i årskurs 3". Rapport, Svenskt Näringsliv.

Heller-Sahlgren, Gabriel och Henrik Jordahl. 2023. "Test scores and economic growth: update and extension." *Applied Economics Letters*, <https://doi.org/10.1080/13504851.2023.2168604>.

Hermes, Henning, Martin Huschens, Franz Rothlauf och Daniel Schunk. 2023. "Motivating low-achievers—Relative performance feedback in primary schools" *Journal of Economic Behavior & Organization* 187: 45–59.

Hift, Richard J. 2014. "Should essays and other 'open-ended'-type questions retain a place in written summative assessment in clinical medicine?" *BMC Medical Education* 14: 249.

Holmlund, Helena, Anna Sjögren och Björn Öckert, 2019, "Jämlikhet i möjligheter och utfall i den svenska skolan," Bilaga 7 till Långtidsutredningen 2019, SOU 2019:40,

Jalava, Nina, Juanna Schrøter Joensen och Elin Pallas. 2015. "Grades and rank: Impacts of non-financial incentives on test performance." *Journal of Economic Behavior & Organization* 115:161–196.

Jewsbury, Paul, Robert Finnigan, Nuo Xi, Yue Jia, Keith Rust och Samantha Burg. 2020. "2017 NAEP Transition to Digitally Based Assessments in Mathematics and Reading at Grades 4 and 8: Mode Evaluation Study." White Paper, National Center for Education Statistics, Institute of Education Sciences, Washington, DC.

Jonsson, Bert, Carola Wiklund-Hörnqvist, Tova Stenlund, Micael Andersson och Lars Nyberg. 2021. "A learning method for all: The testing effect is independent of cognitive ability." *Journal of Educational Psychology* 113(5): 972–985.

Jürges, Hendrik och Kerstin Schneider. 2010. "Central exit examinations increase performance... but take the fun out of mathematics." *Journal of Population Economics* 23 (2): 497–517.

Ladd, Helen F. och Douglas L. Lauen. 2010. "Status versus Growth: The Distributional Effects of School Accountability Policies." *Journal of Policy Analysis and Management* 29(3): 426–450.

Lauen, Douglas L., Gary T. Henry, Roedrick A. Rose och Karen P. Kozlowski. 2013. "The Distribution of Teacher Value Added in North Carolina." Rapport, Consortium for Educational Research and Evaluation, North Carolina.

Lazendic, Goran, Julie-Anne Justus och Stanley Rabinowitz. 2018. "NAPLAN Online Automated Scoring Research Program: Research Report." Rapport, National Assessment Program, <https://nap.edu.au/docs/default-source/default-document-library/naplan-online-aes-research-report-final.pdf>.

Leschnig, Lisa, Guido Schwerdt och Katarina Zigova. 2022. "Central exams and adult skills: Evidence from PIAAC" *Economics of Education Review* 90: 102289.

Ljung, Bengt-Olov. 2000. "Standardproven – 53 år i skolans tjänst". Rapport, PRIM-gruppen, Lärarhögskolan i Stockholm.

Macartney, Hugh. 2016. "The Dynamic Effects of Educational Accountability." *Journal of Labor Economics* 34(1):1–28.

NAP. 2024. "Tailored tests." Hemsida, National Assessment Program, <https://www.nap.edu.au/naplan/understanding-online-assessment/tailored-tests>.

NCES. 2005. "Important Aspects of No Child Left Behind Relevant to NAEP." Hemsida, National Center for Educational Progress, Washington DC, <https://nces.ed.gov/nationsreportcard/nclb.aspx#:~:text=Topics%20for%20the%20Long%20Term,the%20science%20and%20writing%20assessments>.

NCES. 2024. "Assessment Calendar." Hemsida, National Center for Educational Progress, Washington DC, <https://nces.ed.gov/nationsreportcard/about/calendar.aspx>.

Newton, Paul E. 2007. "Clarifying the purposes of educational assessment" *Assessment in Education: Principles, Policy & Practice* 14: 149–170.

Newton, Paul E. 2023. "Maintaining standards: During normal times and when qualifications are reformed." Rapport, Ofqual.

NFER. 2023. "National Reference Test". Hemsida, <https://www.nfer.ac.uk/for-schools/participate-in-research/participate-in-research-projects/national-reference-test/>.

Nusche, Deborah, Thomas Radinger, Torberg Falch och Bruce Shaw. 2016. OECD Reviews of School Resources: Denmark 2016. Paris: OECD Publishing.

OECD. 2023. "PISA 2022 Technical Report". Rapport, OECD Publishing, Paris.

Ofqual. 2018. "Improving GCSE and A level standard setting: a world of learning". Rapport, <https://www.gov.uk/government/publications/improving-gcse-and-a-level-standard-setting-a-world-of-learning>.

Ofqual. 2023. "Background report: National Reference Test Information". Rapport, <https://www.gov.uk/government/publications/national-reference-test-your-information-and-your-privacy/background-report-national-reference-test-information>.

Ofsted. 2024. "School inspection handbook". Manual, <https://www.gov.uk/government/publications/school-inspection-handbook-eif/school-inspection-handbook-for-september-2023>.

Prop. 2017/18:14. "Nationella prov – rättvisa, likvärdiga, digitala." Regeringsproposition.

Robinson, Natasha. 2018. "NAPLAN: Robot marking of school tests scrapped by education ministers." Australian Broadcasting Corporation, <https://www.abc.net.au/news/2018-01-29/push-to-have-robots-mark-naplan-tests-scrapped/9370318>.

Ramesh, Dadi och Suresh Kumar Sanampudi. 2022. "An automated essay scoring systems: a systematic literature review." *Artificial Intelligence Review* 55: 2495–2527.

Roedinger III, Henry L, Adam L. Putnam och Megan A. Smith. 2011. "Ten Benefits of Testing and Their Applications to Educational Practice." *Psychology of Learning and Motivation* 55: 1–36.

Sireci, Stephen G., Thanos Patelis, Saba Rizavi, Alan M. Dillingham och Georgette Rodriguez. 2000. "Setting Standards on a Computerized-Adaptive Placement Examination". Conference Paper, the Annual Meeting of the National Council on Measurement in Education.

Skolinspektionen. 2021. "Ombedömning av nationella prov 2019". Dnr: 2019:503, Stockholm.

Skolverket. 2007. "Hur går det för eleverna i årskurs 5 på de nationella proven?". Rapport, Stockholm.

Skolverket. 2008. "Central rättning av nationella prov". Rapport, Stockholm.

Skolverket. 2010a. "Ämnesproven 2010 i grundskolans årskurs 5". Rapport, Stockholm.

Skolverket. 2010b. "Adaptiva och andra datorbaserade prov - en kunskapsöversikt." Stockholm.

Skolverket. 2015. "Value added." PM, Stockholm.

Skolverket. 2017. "Skolverkets systemramverk för nationella prov." Rapport, Stockholm.

Skolverket. 2019. "Analyser av likvärdig betygssättning mellan elevgrupper och skolor." Rapport 475, Stockholm.

Skolverket. 2020. "Likvärdiga betyg och meritvärden - ett kunskapsunderlag om modeller för att främja betygens och meritvärdens likvärdighet." Rapport 2020:7, Stockholm.

Skolverket. 2022. "Redovisning av uppdrag om att införa central rättning av nationella prov." Dnr:2021:1559, Stockholm.

Skolverket. 2024. "Antagning till gymnasieskolan", <https://www.skolverket.se/regler-och-ansvar/ansvar-i-skolfragor/antagning-till-gymnasieskolan>.

SOU 2014:12. "Utvärdera för utveckling - om utvärdering av skolpolitiska reformer". Statens offentliga utredningar, Stockholm.

SOU 2016:25. "Likvärdigt, rättssäkert och effektivt - ett nytt nationellt system för kunskapsbedömning". Statens offentliga utredningar, Stockholm.

STA. 2017. "Assessment framework for the development of the Year 1 phonics screening check". Rapport, Standards & Testing Agency, London.

STA. 2020a. "Assessment framework: Reception Baseline Assessment". Rapport, Standards & Testing Agency, London.

STA. 2020b. "2019 national curriculum test handbook". Rapport, Standards & Testing Agency, London.

STA. 2023a. "Key stage 2 assessment and reporting arrangements". Rapport, Standards & Testing Agency, London.

STA. 2023b. "Multiplication tables check: administrative guidance". Rapport, Standards & Testing Agency, London.

Stullich, Stephanie, Elizabeth Eisner och Joseph McCrary. 2007. "National Assessment of Title 1: Final Report. Volume I: Implementation". Rapport, Institute of Education Sciences, Washington DC.

Stringer, Neil S. 2012. "Setting and maintaining GCSE and GCE grading standards: the case for contextualised cohort-referencing." Research Papers in Education 27(5): 535-554.

Tåhlin, Michael 2007. "Överutbildningen i Sverige – utveckling och konsekvenser." I Johan Olofsson (red.) Utbildningsvägen – vart leder den? Om ungdomar, yrkesutbildning och försörjning. Stockholm: SNS Förlag.

UVM. 2023a. "Datadokumentation for Testresultater." Rapport, Børne- og Undervisningsministeriet, https://uddannelsesstatistik.dk/Documents/Grundskole/Datadokumentation/DVH_Grundskole_Nationale%20test.pdf.

UVM. 2023b. "Examinations and Other Forms of Assessment." Hemsida, Børne- og Undervisningsministeriet, <https://eng.uvm.dk/primary-and-lower-secondary-education/the-folkeskole/examinations-and-other-forms-of-assessment>.

UVM. 2023c. "Anvendelse af 7-trins-skalaen." Hemsida, Børne- og Undervisningsministeriet, <https://www.uvm.dk/uddannelsessystemet/7-trins-skalaen/anvendelse-af-7-trins-skalaen>.

UVM. 2023d. "Frie grundskoler uden prøver." Hemsida, Børne- og Undervisningsministeriet, <https://www.uvm.dk/frie-grundskoler/proever-test-og-evaluering/skoler-uden-proever>.

Vlachos, Jonas. 2023. "Mervärden i skolan – värdefulla men begränsa(n)de." Ekonomisk Debatt 41(1): 6-18.

Walsh, Elias och Eric Isenberg. 2015. "How Does Value Added Compare to Student Growth Percentiles?" Statistics and Public Policy 2(1):1-13.

Weiss, David J. och G. Gage Kingsbury. 1984. "Application of computerized adaptive testing to educational problems." Journal of Educational Measurement 21(4):361-375.

Wheadon, Chris. 2023. "Would you ask GPT-4 to mark your essays?" Blogg, No More Marking, <https://substack.nomoremarking.com/p/would-you-ask-gpt-4-to-mark-your>.

Wiberg, Marie. 2021. "Befintliga och framtida provmodeller i vuxenproven". Bilaga 1 "Redovisning av uppdrag om nationella prov inom kommunal vuxenutbildning", Skolverket, Stockholm.

Wikström, Christina och Anna Lind Pantzare. 2018. "Standard setting in Sweden: school grades and national tests." I Jo-Anne Baird, Tina Isaacs, Dennis Opposs och Lena Gray (red.). Examination standards: how measures and meanings differ around the world. London: UCL Press.

Näringslivets skolforum är ett initiativ från Svenskt Näringsliv för att stärka Sveriges kompetensförsörjning och förbättra kunskapsresultaten i svensk skola. Syftet är att erbjuda en arena för ökad probleminsikt, förutsättningslös dialog, internationell utblick och erfarenhetsutbyte. svensknaringsliv.se/skolforum

Om författaren



Foto: Karl Gabor

Gabriel Heller-Sahlgren är forskare vid Institutet för Näringslivsforskning och London School of Economics med fokus på utbildningsekonomi. Han är bland annat författare till böckerna Glädjeparadoxen – Historien om skolans uppgång, fall och möjliga upprättelse (tillsammans med Nima Sanandaji, Dialogos, 2019) och Real Finnish Lessons: The True Story of an Education Superpower (CPS, 2015). Han har masters- och doktorsexamina i social policy från London School of Economics och en kandidatexamen i social and political sciences från University of Cambridge.



Att mäta kunskaper och skapa information

Maj 2024, Näringslivets skolforum, Stockholm