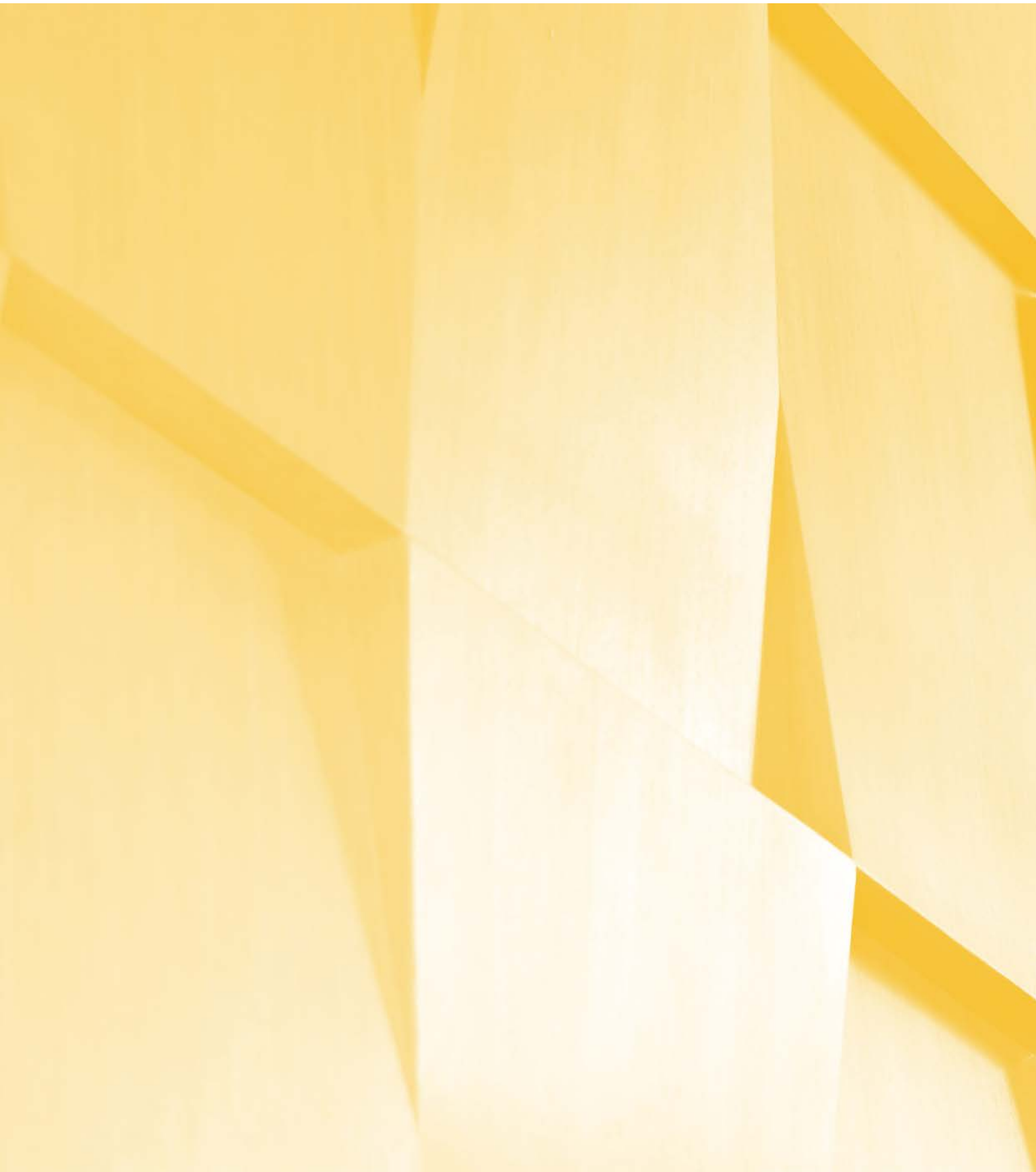


Förädlingsvärdets värde – att mäta skolors och lärares effektivitet



Författare: Gabriel Heller-Sahlgren
London School of Economics
Institutet för Näringslivsforskning

Sammanfattning

Ett effektivt skolsystem kräver rigorösa kunskapsmätt samt information som skapas utifrån dessa mått. Relevanta kvalitetsmätt behövs för att utvärdera både skolor och lärare samt för att stödja deras pedagogiska arbete.

Den här rapporten diskuterar vikten av kvalitetsmätt som baseras på förädlingsvärden – skolors och lärares bidrag till elevernas inläring och andra relevanta utfall – samt hur sådana mått skulle kunna utformas i den svenska skolan.

Forskningen 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. Samtidigt indikerar forskning att lärare påverkar viktiga utfall bland eleverna som inte nödvändigtvis fångas upp i förädlingsvärden som baseras på provresultat. Detta indikerar att det är viktigt att mäta lärares och skolors effektivitet både på prov och enligt andra utfall.

Tyvärr har det svenska utbildningssystemet länge saknat rigorös kvalitetsinformation och fortsätter att göra det i dag. Detta skiljer sig från länder som Norge, England och USA, där man mäter och ofta även publicerar skolors förädlingsvärden. I USA mäter flera delstater och skoldistrikt även enskilda lärares förädlingsvärden.

Rapporten diskuterar vad som kan göras inom ramen för dagens system och de data som finns att tillgå, och exemplifierar med data från nationella prov i årskurs 6 och 9 från en svensk kommun. Analysen visar att modellen för det enda justerade kvalitetsmätt som existerar nationellt – Salsa – tycks fånga upp högstadieskolors kvalitet i ganska hög utsträckning, men att den också missar viktiga skillnader. Ett oroväckande resultat av analysen är samtidigt att man inte kan beräkna ett förädlingsvärde för en hög andel elever, på grund av bortfall i provstatistiken.

Rapporten ger en rad förslag på vad som bör göras för att kunna skapa rigorösa mått på förädlingsvärden i det svenska skolsystemet:

1. **Rätta de nationella proven externt och minska möjligheterna till fusk**
Utan extern och anonym rättning blir det svårt att skapa ett rättssäkert system för att kunna mäta skolors förädlingsvärden på nationella prov. För att minimera sannolikheten att fusk förekommer bör proven hanteras digitalt av Skolverket, medan provskrivningen bör övervakas av oberoende skrivvakter. Alla prov bör också skrivas samtidigt över hela landet.
2. **Säkerställ att de nationella proven skrivs och att resultaten rapporteras in**
Man måste säkerställa att alla elever faktiskt skriver de nationella proven. Myndigheter måste också följa upp så att alla elevers resultat rapporteras in korrekt. Detta förfarande bör underlättas av att hanteringen av proven sker digitalt.
3. **Säkerställ att man kan beräkna förädlingsvärden från första klass**
För att kunna mäta förädlingsvärden från första klass bör elever skriva ett diagnostiskt prov precis innan de börjar skolan – eller precis efter att de har börjat – och resultaten från detta prov bör samlas in.

4. Skapa mått som baseras på årliga diagnostiska prov med relativ rättning

Med hjälp av årliga normrelaterade prov förbättras möjligheterna till uppföljning och insatser – liksom förädlingsvärdenas stabilitet – samtidigt som man kan förändra provformatet, för att på så sätt minska möjligheterna för lärare och skolor att endast coacha elever inför prov snarare än att förbättra deras faktiska kunskaper.

5. Mät förädlingsvärden som baseras på andra utfall än endast prov

För att kunna skapa mått på förädlingsvärden vad gäller icke-kognitiva utfall krävs att data för sådana utfall – såsom frånvaro – börjar rapporteras in till myndigheterna. Ett enklare alternativ vore dock att skapa förädlingsvärden som baseras på elevers genomsnittliga betyg, vilket forskningen indikerar fångar upp icke-kognitiva förädlingsvärden relativt väl. Dessa betyg bör samtidigt förankras i de nationella proven som rättas externt, så att skolornas medelbetyg inte får överstiga deras snittresultat på dessa prov.

6. Samla in alla tillgängliga betyg i slutet av varje läsår

Den norska motsvarigheten till Skolverket beräknar gymnasieskolors förädlingsvärden i varje årskurs för olika program. För att kunna beräkna liknande värden i Sverige inom rimlig tid krävs att man börjar samla in dessa data årligen. Om man även börjar samla in betygen i årskurs 7 och 8 i grundskolan med samma tidsintervall skulle man också kunna mäta förädlingsvärden baserade på betyg årligen under högstadiet.

7. Säkerställ att man kan koppla lärare till elever i nationella databaser

Man kan i dag inte mäta lärares effektivitet på nationell nivå, eftersom det inte går att koppla elever till de lärare som har undervisat dem. För att man ska kunna beräkna enskilda lärares förädlingsvärden krävs att man börjar samla in denna information systematiskt.

8. Tillsätt en expertgrupp för att utreda exakt hur måtten bör utformas

Genom att jämföra i vilken utsträckning olika mått fångar upp lärares och skolors effektivitet, deras prediktiva värde för långsiktiga utfall samt vilka möjligheter till manipulation de skapar skulle man maximera möjligheterna att skapa så rigorösa mått som möjligt. Man bör exempelvis utreda huruvida svenska lärare och skolor är olika effektiva för olika slags elever. Om så är fallet bör måtten som tas fram också reflektera detta.

Innehåll

Sammanfattning	1
1. Introduktion.....	4
2. Kvalitetsinformation påverkar skolsystemet	7
3. Förädlingsvärdets värde.....	10
4. Effekter av informationssystem baserade på förädlingsvärden	14
5. Svenska mått på utbildningskvalitet och deras brister	18
6. Andra länders mått på förädlingsvärden.....	22
6.1 Norge.....	22
6.2 England.....	23
6.3 USA.....	24
7. Hur kan man mäta förädlingsvärden i Sveriges skolsystem?.....	26
7.1 Förädlingsvärden från en svensk kommun	27
8. Förslag på hur man skapar mer rigorösa mått på förädlingsvärden.....	33
8.1 Rätta de nationella proven externt och minska möjligheterna till fusk	33
8.2 Säkerställ att de nationella proven skrivs och att resultaten rapporteras in.....	33
8.3 Säkerställ att man kan beräkna förädlingsvärden redan från första klass	34
8.4 Skapa mått som baseras på årliga diagnostiska prov med relativ rättning	34
8.5 Mät förädlingsvärden som baseras på andra utfall än endast prov	35
8.6 Samla in alla tillgängliga betyg i slutet av varje läsår	35
8.7 Säkerställ att man kan koppla lärare till elever i nationella databaser	35
8.8 Tillsätt en expertgrupp för att utreda exakt hur måtten bör utformas.....	36
9. Slutsatser.....	37
Referenser.....	39

1. Introduktion

Under de senaste decennierna har kunskapskraven på den svenska arbetsmarknaden ökat markant, i takt med att den teknologiska utvecklingen har rationaliserat bort många av de enkla jobb som tidigare existerade (Tåhlin 2007). Detta har gjort att i ett internationellt perspektiv är betydelsen av kunskaper och färdigheter stor för sannolikheten att ha ett heltidsarbete i dag i Sverige (Heller-Sahlgren 2019). Sammantaget har den ekonomiska avkastningen på både kognitiva och icke-kognitiva färdigheter ökat avsevärt i Sverige sedan tidigt 1990-tal (Edin m.fl. 2017). Det mesta tyder också på att utvecklingen kommer att fortsätta i samma riktning (Cedefop 2015). För att säkra framtidens kompetensförsörjning och tillväxt behövs därför ett större fokus på kunskaper och andra relevanta utfall i skolsystemet.

Ett effektivt skolsystem kräver i sin tur rigorösa kunskapsmått samt information som skapas utifrån dessa mått. En välfungerande skolmarknad bygger exempelvis på att föräldrar och elever kan göra informerade val – samt att kvalitetsmåten som lärare och skolor bedöms efter ger rätt incitament för dessa aktörer. För att myndigheter och huvudmän ska kunna identifiera och åtgärda problem i skolsystemet måste de också veta var problemen finns, samt vilka delar som fungerar väl (Heller-Sahlgren och Jordahl 2016; OECD 2008). Relevanta kvalitetsmått behövs därför för att utvärdera både skolor och lärare samt för att stödja deras pedagogiska arbete.

Den här rapporten diskuterar vikten av kvalitetsmått som baseras på förädlingsvärden – skolors och lärares bidrag till elevernas inläring och andra relevanta utfall – samt hur sådana mått skulle kunna utformas i den svenska skolan. De senaste decennierna har ett antal länder skapat system för att mäta både skolors och lärares förädlingsvärden. Det är viktigt att studera i vilken utsträckning Sverige kan lära av forskningen från dessa system.

Forskningen 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 skolors och lärares förädlingsvärden har ofta också positiva effekter på inläringen. Samtidigt indikerar forskning att lärare påverkar viktiga utfall bland eleverna som inte nödvändigtvis fångas upp i förädlingsvärden som baseras på provresultat. Detta indikerar att det är viktigt att mäta lärares och skolors effektivitet både på prov och enligt andra utfall.

Tyvärr har det svenska utbildningssystemet länge saknat mått på förädlingsvärden inom skolsystemet – och de kvalitetsmått som existerar är inte tillräckligt robusta. Detta skiljer sig från länder som Norge, England och USA, där man i dag mäter och ofta även publicerar skolors förädlingsvärden. I USA mäter flera delstater och skoldistrikt även enskilda lärares förädlingsvärden.

Inom ramen för dagens svenska system kan man skapa förädlingsvärden på skolnivå från och med årskurs 3, då elever för första gången skriver nationella prov i matematik och svenska. I årskurs 6 skriver elever sedan för första gången nationella prov i alla tre kärnämnen och får betyg i en rad ämnen, som dock varierar mellan olika skolor. Rapporten exemplifierar hur man skulle kunna gå tillväga för att skapa mått på

skolors förädlingsvärden vad gäller resultaten i nationella prov i högstadiet, med data på elevnivå från en svensk kommun. Analysen visar att modellen för det enda justerade kvalitetsmått som existerar nationellt – Salsa – tycks fånga upp högstadieskolors kvalitet i ganska hög utsträckning, men att den också missar viktiga skillnader. Ett oroväckande resultat är samtidigt att man inte kan beräkna ett förädlingsvärde för en förvånansvärt hög andel elever, på grund av bortfall i provstatistiken.

Rapporten ger till sist en rad förslag på hur man kan skapa bättre kvalitets- och informationsmått i Sverige. Utan extern och anonym rättning blir det svårt att skapa ett rättssäkert system för att kunna mäta skolors förädlingsvärden på nationella prov. För att minimera sannolikheten att fusk förekommer bör proven hanteras digitalt av Skolverket, medan provskrivningen bör övervakas av oberoende skrivvakter. Alla prov bör också skrivas samtidigt över hela landet. Likaså måste man i bättre utsträckning än i dag säkerställa att alla obligatoriska prov faktiskt skrivs av eleverna och att resultaten rapporteras in korrekt. Detta förfarande bör underlättas av att hanteringen av proven sker digitalt.

För att kunna mäta förädlingsvärden redan från första klass bör elever skriva ett diagnostiskt prov precis innan de börjar skolan – eller precis efter att de har börjat – och resultaten från detta prov bör samlas in. Ett mer radikalt förslag vore dock att införa nya externt rättade och normrelaterade diagnostiska prov som genomförs årligen från första klass. Med hjälp av sådana prov förbättras möjligheterna till uppföljning och insatser – liksom förädlingsvärdenas stabilitet – samtidigt som man kan förändra provformatet, för att på så sätt minska möjligheterna för lärare och skolor att endast coacha elever inför prov snarare än att förbättra deras faktiska kunskaper.

För att kunna skapa mått på förädlingsvärden vad gäller bredare utfall krävs samtidigt att data för sådana utfall – såsom frånvaro – börjar rapporteras in till myndigheterna. Ett enklare alternativ vore dock att skapa förädlingsvärden som baseras på elevers genomsnittliga betyg, vilket forskningen indikerar fångar upp icke-kognitiva förädlingsvärden relativt väl. Dessa betyg bör samtidigt förankras i de nationella proven som rättas externt, så att skolornas medelbetyg inte får överstiga deras snittresultat på dessa prov.

I dag samlas betygen i årskurs 7 och 8 inte in över huvud taget, medan kursbetygen i årskurs 1 och 2 på gymnasiet endast samlas in i slutet av gymnasieutbildningen. Den norska motsvarigheten till Skolverket beräknar i dag gymnasieskolors förädlingsvärden i varje årskurs för olika program. För att kunna beräkna liknande värden i Sverige inom rimlig tid krävs att man börjar samla in dessa data årligen. Om man även börjar samla in betygen i årskurs 7 och 8 i grundskolan med samma tidsintervall skulle man också kunna mäta förädlingsvärden vad gäller betyg årligen på högstadiet.

Det är dock viktigt att inte endast fokusera på skolors förädlingsvärden. Den största variationen i lärares förädlingsvärden existerar inom skolor; att mäta enskilda lärares effektivitet är också viktigt. Man bör därför också samla in information som gör det möjligt att koppla elever till lärare på nationell nivå.

Mer generellt vore det önskvärt att utreda exakt hur måtten på förädlingsvärden bör utformas, för att säkerställa att de fångar upp skolors och lärares bidrag till elevers utfall på ett tillfredsställande sätt i den svenska kontexten. Man bör exempelvis utreda huruvida svenska lärare och skolor är olika effektiva för olika slags elever. Om så är fallet bör måtten som tas fram också reflektera detta.

Rapporten rekommenderar därför att regeringen tillsätter en expertgrupp av forskare för att utforma och analysera olika mått på skolors och lärares förädlingsvärden, både vad gäller provresultat och bredare utfall. Genom att jämföra i vilken utsträckning dessa olika mått fångar upp lärares och skolors effektivitet, deras prediktiva värde för långsiktiga utfall samt vilka möjligheter till manipulation de skapar skulle man maximera möjligheterna att skapa så rigorösa mått som möjligt.

2. Kvalitetsinformation påverkar skolsystemet

Information om skolors och lärares kvalitet är viktig för alla aktörer inom utbildnings-systemet. För att föräldrar och elever ska kunna välja en bra skola är tillgången till informationen naturligtvis av stor vikt; om föräldrarna inte vet vilken kvalitet skolor håller blir det svårt att göra ett välinformerat val (Heller-Sahlgren och Jordahl 2016). Dessa val görs inom ramen för en ”kvasimarknad”, där skattebetalarna som en tredje part finansierar skolpengen. Normalt sett aggregerar marknader svåråtkomlig och lokal information via prismekanismen. Priser ger lättförståeliga signaler till producenter och konsumenter om hur de ska agera (Hayek 1945). Men eftersom skolpengen helt och hållet finansieras offentligt existerar dessa signaler inte i skolvalssystemet. Dessutom har skolor ett informationsövertag gentemot föräldrar, vilket gör att de kan bygga ett bra rykte utan att nödvändigtvis höja kvaliteten så länge informationen inte är tillräcklig (MacLeod och Urquiola 2012). Internationell forskning tyder också på att konsekvenserna av felaktiga val kan bli stora och att bristfällig information kan bidra till att föräldrar gör sämre val (Ajayi 2012; Lai, Sadoulet och de Janvry, 2009; Lucas och Mbiti 2012), medan föräldrars och elevers skolval ofta påverkas kraftigt av interventioner som ökar tillgången till kvalitetsinformation (se Ajayi m.fl. 2017; Andrabi m.fl. 2017; Autor m.fl. 2016; Hastings och Weinstein 2018; Gallego et al. 2012). En färsk randomiserad studie från Linköping indikerar att detta även gäller i en svensk kontext: när föräldrar ges information om skolors kvalitet börjar de i genomsnitt välja bättre presterande skolor (Kessel och Olme 2018). För att skolvalet ska fungera så effektivt som möjligt behöver föräldrar och elever alltså tillgång till så god information som möjligt om skolors kvalitet.

Likaså är det A och O att myndigheter som Skolinspektionen har tillgång till relevant information för att kunna identifiera vilka skolor och lärare som är effektiva och vilka som är ineffektiva. Utan rätt information kan myndigheterna inte stärka skolors incitament att förbättra sin kvalitet – eller avgöra om skolor bör stängas på grund av att de håller låg kvalitet. Kvalitetsinformation är också av vikt för skolledning och huvudmän, framför allt för att kunna identifiera svagheter och styrkor bland skolor och lärare (både existerande och de som söker jobb på skolorna). Detta är avgörande eftersom det har visat sig vara svårt att bedöma lärares effektivitet med hjälp av deras observerbara egenskaper, exempelvis vilken slags utbildning de har (Jackson m.fl. 2014). Inte heller har man kunnat bedöma deras effektivitet genom att observera och bedöma dem. Forskning indikerar nämligen att elevenkäter och rektorsutvärderingar endast kan förklara ett par procentenheter av skillnaden i lärares kapacitet att förbättra elevers kunskaper (se Jacob och Lefgren 2008; Kane m.fl. 2013; Rockoff m.fl. 2012; Strong m.fl. 2011). Trots att de flesta rektorer tror att de vet vilka lärare som är effektiva och vilka som är ineffektiva kan deras omdömen oftast inte förklara mer än en bråkdel av de faktiska skillnader som existerar. Det finns ingen anledning att tro att det skulle vara lättare att identifiera bra och dåliga skolor med blotta ögat, exempelvis genom skolinspektioner. De flesta av oss tror självfallet att vi kan separera en bra eller dålig skola eller lärare när vi ser dem. Men detta stämmer ofta inte.

Det finns flera möjliga anledningar till att så är fallet. En är att de flesta har preferenser för specifika skol- och inlärningsmiljöer som man omedvetet letar efter under observationer av skolor och lärare, trots att dessa inte nödvändigtvis hjälper elever

att lära sig. Mycket av den inläring som sker är också osynlig, vilket betyder att saker som de flesta associerar med fungerande skolverksamhet per automatik inte betyder att eleverna lär sig något; elever kan exempelvis vara upptagna med skolarbete och engagerade i klassrum som karaktäriseras av lugn och ro, utan att de lär sig någonting för det. Det finns dessutom psykologisk forskning som tyder på att människor missar extremt mycket information när de observerar saker mer generellt – ett fenomen som kallas ”uppmärksamhetsblindhet” – vilket naturligtvis har implikationer för vår förmåga att observera skolors och lärares relativa effektivitet (se Coe 2014). Dessa är endast ett par tänkbara förklaringar till varför vi är dåliga på att identifiera skolors och lärares kvalitet med blotta ögat.

För att kunna bedriva ett systematiskt kvalitetsarbete bör aktörer inom skolsystemet alltså ha direkt tillgång till relevanta data som fångar upp skolors och lärares bidrag till inläringen; rigorös kvalitetsinformation i skolsystemet behövs för att utvärdera skolor och lärare samt för att stödja det pedagogiska arbetet överlag.

Empirisk forskning tyder mycket riktigt på att tillgången till kvalitetsinformation i utbildningssystemet – skolval eller ej – påverkar skolors incitament avsevärt, antingen direkt eller via de incitament som skapas av att myndigheter håller skolor ansvariga för deras prestationer. Forskning från olika amerikanska delstater under 1990-talet visar exempelvis att offentliggöranden av skolors kvalitet tillsammans med konsekvenser för skolors prestationer ledde till förbättringar (Hanushek och Raymond 2005). En rad andra amerikanska studier finner liknande effekter av att myndigheter håller skolor ansvariga med hjälp av information om deras kvalitet (Ballou och Springer 2017; Carlson m.fl. 2013; Chakrabarti 2014; Reback m.fl. 2014; Winters m.fl. 2010; Winters och Cowen 2012; Witte m.fl. 2014).¹ Detsamma gäller studier från en rad andra länder: mer information i systemet förbättrar generellt skolors resultat, speciellt bland skolor som har starkare incitament på grund av konkurrens (se till exempel Andrabi m.fl. 2017; Burgess m.fl. 2013; Camargo m.fl. 2018; Koning och van der Wiel 2012).

De flesta studier fokuserar på elevprestationer i ett kortare perspektiv. Deming m.fl. (2016) finner dock att ett system för ansvarsutkrävande via information om skolors prestationer som fanns i Texas under 1990-talet påverkade elevers långsiktiga resultat. Skolor som var nära att bli bedömda som underpresterande förbättrade sina resultat på prov som knöts till systemets incitament. Som ett resultat av detta ökade sannolikheten att dessa skolors elever läste vidare och slutförde en examen på universitet, och att lågpresterande elever i dessa skolor även hade högre löner vid 25 års ålder. Samtidigt finner författarna inga effekter bland skolor som var nära att bli bedömda som ”bra”, men att lågpresterande elever i dessa skolor i stället når sämre resultat och har lägre löner senare i livet. Dessa skolor började klassificera fler elever som särskoleelever – som inte omfattas av kvalitetsinformationen – vilket i sin tur helt och hållet förklarar de negativa effekterna. Även om effekten av ansvarsutkrävande via kvalitetsinformation i genomsnitt alltså var positiv, och det var lågpresterande elever i lågpresterande skolor som tjänade mest, står det klart att lågpresterande elever som gick i bättre skolor förlorade på det.

Detta visar tydligt att det finns oavsiktliga konsekvenser som måste tas hänsyn till i utformningen av systemet för att maximera dess positiva effekter. All kvalitetsinformation är naturligtvis inte lika bra – och all information bör inte ens nödvändigtvis offentliggöras. Beroende på vilken information man fokuserar på förändras helt

¹ Likaså tyder amerikansk forskning på att sådana system minskar andelen frånvarande lärare i systemet, effekter som drivs av att de minst effektiva lärarna börjar dyka upp i högre utsträckning (Gershenson 2016a).

enkelt incitamentsstrukturen i skolsystemet. Om informationen fokuserar på andelen elever som når godkänt kommer skolor sannolikt att fokusera på dessa elever. Om man inte inkluderar alla elever i den information som presenteras finns det också en risk för att skolor försöker exkludera de som presterar dåligt. Många system har också skapat incitament att manipulera resultaten, till exempel genom att öka antalet adhd-diagnoser eller genom att skolorna helt enkelt fokuserar på sådant som mäts i provet men negligerar andra viktiga kunskaper (se Figlio och Loeb 2011). Ett exempel är att skolor kan frestas att placera de bästa lärarna i de årskurser som skriver de prov som omfattas av systemet för ansvarsutkrävande (Fuller och Ladd 2012). Det är dock värt att notera att de studier som diskuteras ovan faktiskt analyserar dåligt utformade system. De positiva effekter som ofta återfinns i forskningen kan inte heller enbart förklaras av manipulation (Heller-Sahlgren och Jordahl 2016).

Trots detta står det hur som helst klart att system som på olika sätt fokuserar på absoluta provresultat – inklusive andelen som når godkänt och genomsnittliga resultat – är långt ifrån tillfredsställande. Det är lätt att tro att bra eller dåliga resultat absolut sett reflekterar skolors och lärares effektivitet. Men detta stämmer inte. I stället reflekterar genomsnittliga provresultat framför allt elevsammansättningen, vilket skolor och lärare naturligtvis inte bör jämföras efter. Om så sker skapas starka incitament bland skolor att fokusera på att locka till sig högrepresterande elever i stället för att öka deras kunskaper och färdigheter. Det blir också svårt för myndigheter och huvudmän att använda informationen för att identifiera styrkor och svagheter i skolors och lärares arbeten. Dålig information om kvalitet gör mer generellt att föräldrar och elever riskerar att välja skolor som inte är effektiva för deras barn (se Heller-Sahlgren och Jordahl 2016). Om kvalitetsmått som skolor och lärare bedöms efter fångar upp saker som de inte kan påverka anses de naturligtvis också med rätta vara orättvisa, vilket underminerar både deras legitimitet och användbarhet (OECD 2008). Allt detta tydliggör vikten av att kvalitetsinformationen utformas på rätt sätt. Det är här skolors och lärares ”förädlingsvärden” – ett mått på skolors och lärares faktiska bidrag till inläringen – kommer in i bilden. Nästa avsnitt diskuterar förädlingsvärdets fördelar som kvalitetsmått.

3. Förädlingsvärdets värde

Med förädlingsvärdet menas den förändring i elevers prestationer som kan tillskrivas skolor eller enskilda lärare, efter att andra faktorer har rensats bort. Till skillnad från absoluta mått på elevers utfall är det meningen att skolors och lärares förädlingsvärden ska fånga upp deras bidrag till dessa utfall, vilket i sin tur antas vara ett rättvist mått på deras effektivitet. I grunden är vi huvudsakligen intresserade av skolors och lärares bidrag till elevers humankapital, vilket kan fångas upp av olika mått. Det vanligaste är att man analyserar provresultat för att mäta förädlingsvärden, helt enkelt för att detta ofta är de enda objektiva utfall man har att tillgå i utbildningssystemet. Man antar också att kvalitetsmått som baseras på elevers provresultat är relevanta, eftersom många studier kopplar dessa resultat till deras utfall senare i livet (se Jackson m.fl. 2014).

Med information om förädlingsvärden får skolor och lärare större möjligheter att bedriva ett systematiskt, datadrivet kvalitetsarbete. Likaså blir det lättare för myndigheter att bedöma skolors effektivitet, vilket är avgörande för att kunna ta reda på vilka interventioner som fungerar (OECD 2008). Information som skapas efter skolors och lärares förädlingsvärden ger också bättre incitament bland aktörerna att höja utbildningskvaliteten, eftersom föräldrar och myndigheter i högre utsträckning kan separera effektiva skolor från ineffektiva skolor. Medan system som baseras på förädlingsvärden ökar pressen på skolor att öka förädlingsvärdet, skapar system som baseras på provresultat högre press på skolor att nå högre absoluta provresultat. Alla möjligheter att manipulera resultaten som skolorna och lärarna bedöms efter försvinner nödvändigtvis inte (se till exempel Macartney 2016), men detta beror också i hög utsträckning på hur systemet är utformat samt hur det används.

I praktiken kan en skolas eller lärares förädlingsvärde beräknas på olika sätt, varav det enklaste sättet är att subtrahera elevernas standardiserade provresultat, innan de börjar på skolan eller börjar undervisas av en enskild lärare, från deras standardiserade provresultat efter att de har lämnat skolan eller inte längre undervisas av läraren. Med denna metod är det dock svårt att jämföra skolor och lärare, eftersom olika elever presterar på olika nivåer initialt sett. Man brukar därför justera för elevernas initiala resultat innan de började på skolan eller börjar undervisas av läraren ifråga (Koedel m.fl. 2015). På så sätt försöker man rensa bort betydelsen av elevernas initiala resultat för deras progression när man beräknar elevernas förädlingsvärde. I vissa fall tar man även hänsyn till elevers bakgrund i beräkningarna, för att ytterligare maximera sannolikheten att jämföra lika med lika.

Men går det verkligen att på detta sätt skapa tillförlitliga mått på skolors och lärares effektivitet? Denna fråga har utvärderats genom att man jämför det beräknade förädlingsvärdet med effekten av att elever blir slumpmässigt placerade på en specifik skola eller hos en specifik lärare. Denna forskning tyder på att både skolors och lärares förädlingsvärden fångar upp deras effektivitet på ett tillfredsställande sätt så länge man har tillgång till elevers tidigare provresultat (Bacher-Hicks m.fl. 2017; Bifulco 2012; Cullen m.fl. 2017; Deming 2014; Glazerman och Protik 2015; Kane

och Staiger 2008; Kane m.fl. 2013).² Även när forskningen finner att det traditionella förädlingsvärdet inte helt och hållet mäter precis alla skolors effektivitet är dess prediktiva värde högt i genomsnitt – och slutsatsen är att politiska beslut som baseras på förädlingsvärden som justerar för elevers initiala resultat kan ha stora positiva effekter på elevers resultat (se Angrist m.fl. 2016, 2017). Med andra ord tyder studier, som jämför ganska enkelt framräknade förädlingsvärden med den kausala effekten av lärare och skolor i experiment, på att förädlingsvärdena i hög grad är tillförlitliga.

Ett möjligt problem med att jämföra förädlingsvärden med effekter i experiment är att det kan vara svårt att dra starka slutsatser kring den externa validiteten av resultaten; effekterna i den randomiserade forskningen är inte nödvändigtvis överförbara till lärare och skolor som inte omfattas av analysen. Ett alternativ för att studera förädlingsvärdets validitet är att i stället studera lärare som byter skolor eller börjar undervisa andra årskurser i samma skola, vilket gör att elever får lärare med högre eller lägre förädlingsvärden. Sådan forskning drar dock generellt samma slutsatser i olika kontexter: lärares förädlingsvärden fångar upp deras kausala effekter på elevers resultat på ett tillfredsställande sätt (Chetty m.fl. 2014a; Bacher-Hicks m.fl. 2014; Rothstein 2017).³ Det verkar också som att modeller som förutom elevers initiala resultat också inkluderar bakgrundskontroller, i princip inte skiljer sig från modeller som endast inkluderar elevers initiala resultat (Chetty m.fl. 2014a). Dessa studier stödjer alltså slutsatserna i forskningen som jämför förädlingsvärden med de kausala effekterna enligt slumpmässig placering av lärare eller elever.

Men är förädlingsvärdet prediktivt för elevers utfall senare i livet? Om elever undervisas av lärare och går på skolor som har högre förädlingsvärde bör detta ge utslag i deras långsiktiga utbildnings- och arbetsmarknadsutfall. Amerikansk forskning indikerar att så är fallet. Elever med lärare med högt förädlingsvärde går i högre utsträckning vidare till universitet, blir tonårsföräldrar i mindre utsträckning – och tjänar mer som vuxna. Resultaten indikerar att om vi skulle ersätta de 5 procent minst effektiva lärarna med lärare som har ett genomsnittligt förädlingsvärde skulle elevernas livsinkomster öka med över två miljoner kronor per klassrum (Chetty m.fl. 2014b). Det finns hittills ingen liknande forskning som studerar hur prediktivt skolors genomsnittliga förädlingsvärden är för elevers arbetsmarknadsutfall, men två studier finner att skolors förädlingsvärden som beräknas med elevers provresultat förutspår hur bra eleverna presterar senare på högskolan (Hubbard 2018; Totty 2017). Detta indikerar att skolors och lärares förädlingsvärden som baseras på elevers provresultat också förutspår hur bra elever presterar senare i livet. Detta tyder på att de fångar upp faktisk utbildningskvalitet snarare än skolors och lärares förmåga att enbart förbättra elevers resultat i proven som förädlingsvärdena baseras på – vilket är ett vanligt argument mot standardiserade prov mer generellt.

Det är visserligen viktigt att betona att alla ovanstående studier inte analyserar system som knyter väldigt starka incitament direkt till provresultaten, så som man gör inom ramen för offentligt ansvarsutkrävande. Det finns därför ingen garanti för att förädlingsvärdena skulle vara lika prediktiva om man knyter väldigt starka incitament till dem,

² Viss forskning drar något annorlunda slutsatser när de jämför skolors kausala effekter enligt slumpmässiga placeringar av ett litet antal elever med samma skolors förädlingsvärden beräknade efter ett stort antal elever som inte placeras slumpmässigt (Fortson m.fl. 2012). Detta är dock ett otillräckligt tillvägagångssätt för att mäta tillförlitligheten i förädlingsvärden då skolors effektivitet mycket väl kan skilja sig mellan de två urvalen – vilket gör att man i praktiken jämför äpplen med päron. Till stöd för detta finner samma forskning att de kausala effekterna i det slumpmässiga urvalet inte skiljer sig med statistiskt signifikant marginal från förädlingsvärdena i det icke-slumpmässiga urvalet, när man begränsar det senare till elever som är lika varandra vad gäller observerbara egenskaper.

³ Rothstein (2017) menar samtidigt visserligen också att metoden inte fungerar eftersom förändringar i lärares effektivitet på skolorna korrelerar med förändringar i tidigare resultat på samma skola. Annan forskning visar dock att detta förklaras av mekaniska samband och betyder inte att metoden inte fungerar (Chetty m.fl. 2016, 2017).

till exempel om man använder dem som grund för belöningar och straff, då incitamenten att enbart coacha eleverna för proven och andra former av manipulation också ökar. Hubbards (2018) forskning tyder dock på att så inte nödvändigtvis är fallet, då han analyserar förädlingsvärden som skapas på prov som inkluderas i Michigans system för ansvarsutkrävande. Som diskuteras i avsnitt 8.4: om man vill knyta starka incitament till förädlingsvärdena kan det hur som helst vara värt att skapa mått baserat på kunskapskontroller som kan variera över tid, vilket gör det svårare att manipulera resultaten.

Totalt sett visar alltså ett flertal studier att förädlingsvärden som baseras på prov på ett tillfredsställande sätt fångar upp lärares och skolors kunskapsmässiga kvalitet. Samtidigt finner vissa studier att förädlingsvärden inte alltid är stabila över tid (se Koedel 2015), vilket minskar möjligheterna att använda dem för vissa syften. I England har exempelvis forskning visat att få av skolornas förädlingsvärden skiljer sig från det nationella medelvärdet med statistiskt säkerställd marginal, när man använder dagens förädlingsvärden för att förutspå framtida förädlingsvärden. Detta är relevant för möjligheterna att använda måtten vid skolval, eftersom informationen som engelska föräldrar använder när de väljer skola innan barnet fyller elva år baseras på förädlingsvärden för elever som är sju år äldre (se Leckie och Goldstein 2009).⁴ Den huvudsakliga anledningen till instabilitet är inte att skolor håller samma kvalitet eller att deras kvalitet förändras mycket över tid. I stället handlar det om mätfel som alla provresultat lider av i olika utsträckning, vilket kan minskas genom att använda resultat från fler provtillfällen när man beräknar elevers initiala resultat i modellen för förädlingsvärden (Muriel och Smith 2011). Forskning visar också generellt att det prediktiva värdet av förädlingsvärden för framtida förädlingsvärden ökar om man använder data över flera år. Förädlingsvärden som baseras på årliga data är dessutom betydligt mer stabila (Ferrão 2012; Koedel m.fl. 2015). Exakt hur bra ett informationssystem som baseras på förädlingsvärden är beror alltså delvis på hur ofta man använder proven för att skapa dessa värden.

Frågan är dessutom inte huruvida systemet är perfekt eller inte, utan snarare om utbildningssystemet i sig fungerar bättre med än utan information om skolors och lärares förädlingsvärden. Forskning finner ett stort värde i att använda förädlingsvärden för beslutsfattande även när de inte är stabila, exempelvis när det handlar om vilka lärare och skolor som bör finnas kvar i systemet för att maximera elevers resultat (se Koedel m.fl. 2015). Exempelvis finner Chetty m.fl. (2014a, 2014b) att lärares förädlingsvärden är tillräckligt stabila för att endast ett par års data ska vara tillräckliga för att värdena ska vara av stort värde vid exempelvis anställningsbeslut. Det råder därför ingen större tvekan om att förädlingsvärdet är ett väldigt användbart mått, även om det naturligtvis inte är helt perfekt. Exakt hur användbart det är beror också på vilka data man använder för att utforma måtten.

Samtidigt är det viktigt att notera att de flesta mått på skolors och lärares förädlingsvärden som utformats fokuserar på förädlingsvärden som baseras på provresultat. Men det finns också andra färdigheter som har stor betydelse för elevers framtid. Även om dessa färdigheter i viss utsträckning fångas upp av provresultat har de också en viktig oberoende effekt på betyg och långsiktiga utfall (se Kautz m.fl. 2014; Borghans m.fl. 2016). Det är inte heller säkert att skolor och lärare som är effektiva när det gäller att förbättra elevers provresultat nödvändigtvis är effektiva i alla dessa andra avseenden. Forskning indikerar exempelvis att lärare som är effektiva på att förbättra

⁴ Notera att detta gäller förädlingsvärden av äldre snitt. Som diskuteras i avsnitt 6.2 har systemet gjorts om sedan dess, men hittills har ingen analys genomförts vad gäller stabiliteten i dessa mått.

elevs provresultat inte är effektiva på att höja ett index som inkluderar variabler som frånvaro, kursbetyg och huruvida de fick gå om en kurs eller inte.⁵ Lärarnas förädlingsvärden vad gäller detta index har också ett större prediktivt värde vad gäller sannolikheten att elever går ut gymnasiet och andra långsiktiga utfall, jämfört med förädlingsvärden som fångas upp av prov. Genom att kombinera bägge måtten kan man fördubbla det prediktiva värdet vad gäller långsiktiga utfall, jämfört med det provbaserade förädlingsvärdet. Intressant nog tycks förädlingsvärdet som baseras enbart på elevernas kursbetyg vara i princip lika bra på att förutspå långsiktiga utfall som det som baseras på det kombinerade indexet (Jackson 2018). Detta är av vikt eftersom det är lättare att skapa mått från elevens betyg än de andra utfallen i indexet, som sannolikt kan manipuleras i högre utsträckning. Resultaten tyder alltså på att man bör vara öppen för att skapa förädlingsvärden som baseras på både provresultat och betyg, eftersom dessa utfall kan fånga upp olika saker.

Totalt sett visar forskningen därför att man kan mäta lärares och skolors förädlingsvärden på ett tillförlitligt sätt som fångar upp deras faktiska effektivitet. Den visar också att dessa värden tenderar att vara prediktiva för elevernas utfall senare i livet. Nästa avsnitt diskuterar i sin tur forskningen kring effekterna av informationssystem som huvudsakligen eller i stor utsträckning baseras på just förädlingsvärden.

⁵ Annan forskning finner också att lärare som är effektiva på att förbättra provresultaten inte nödvändigtvis är bra på att minska elevernas frånvaro, öka deras glädje, eller förbättra deras uppträdande (Blazar och Kraft 2017; Gershenson 2016b).

4. Effekter av informationssystem baserade på förädlingsvärden

Hur fungerar informationssystem som baseras på skolors och lärares förädlingsvärden? Forskningen som finns tillgänglig indikerar att de fungerar ganska bra. I Florida – där skolor hålls ansvariga för bland annat sina förädlingsvärden – förbättrar skolorna resultaten, både på prov som omfattas av systemet och på prov som inte gör det. Effekten håller i sig flera år, även när eleven har lämnat skolan (Rouse, m.fl. 2013; Chiang 2009). I New York finner man också att effekten av ett underkänt betyg i myndigheternas rankning, som bland annat inkluderar förädlingsvärden, är positiv för de kommande resultaten i engelska och matematik. Resultatförändringen i engelska, men inte matematik, är positiv under det första året – och ökar med tiden i engelska och växer till att bli signifikant också i matematik (Winters och Cowen 2012). Annan forskning från New York har funnit liknande effekter (Rockoff och Turner 2010). Ingenting tyder heller på att skolorna manipulerade resultaten eller att de slutade erbjuda kurser där elever inte skrev prov som myndigheterna räknade med i rankningen.

En viktig poäng med system som använder förädlingsvärde för ansvarsutkrävande är att de kan antas ge incitament att förbättra resultaten också bland elever som ligger långt ifrån gränsen för godkänt och elever som redan klarar kraven. Forskning visar mycket riktigt att elever som redan klarar kraven inte förlorar på skolrankningar när man använder skolors förädlingsvärden som grund för rankningen. Det kan de däremot göra om proven fokuserar på att öka antalet godkända elever (Ladd och Lauen 2010; Lauen och Gaddis 2012). Detta indikerar att informationssystem som ämnar mäta skolors effektivitet kan ha positiva effekter om man utformar systemet väl. Generellt är det mindre lönsamt för skolor att manipulera förädlingsvärdet eftersom en artificiell ökning av elevers resultat ett år gör det svårare att förbättra resultaten för samma elever året efter (Figlio och Loeb 2011). Om aktörernas incitament knyts till förbättringar snarare än absoluta prestationsnivåer blir det därför mindre lockande för skolor att försöka manipulera informationen.

I North Carolina har man exempelvis valt att knyta skolors incitament till deras förädlingsvärden, och man ger bonusar till de skolor som presterar över förväntan enligt statistiska beräkningar. Bonusarna fördelas mellan alla anställda på skolan. Skolor kan alltså inte bestämma själva hur de vill fördela bonusarna. Samtidigt presenteras resultaten offentligt. Forskning finner positiva effekter av att precis missa gränsen för att få en bonus på resultaten året efteråt – i läsförståelse, matematik samt naturvetenskap. Att effekten inte är negativ, utan snarare positiv, i naturvetenskap indikerar att resultatförbättringarna inte skedde på grund av att man ignorerade ämnen utan incitament – resultaten i naturvetenskap spelar ingen roll för huruvida skolor tilldelas bonusar (Lauen 2013). Detta stödjer idén om att systemets utformning spelar stor roll för hur informationssystemen påverkar skolor.

Samtidigt är det viktigt att notera att det självklart fortfarande kan finnas incitament och möjligheter att manipulera förädlingsvärden, speciellt över en längre tidsperiod. Detta eftersom måttet beror på initiala resultat: goda provresultat ett år kan göra det svårare att skapa ett högt förädlingsvärde under nästa period. Eftersom elever endast bidrar till förädlingsvärden så länge de är i skolan kan det därför finnas starkare incitament att förbättra prestationerna under de sista åren i skolan – goda prestationer bland dessa elever spelar mindre roll för skolornas framtida förädlingsvärden, eftersom

de snart lämnar skolan. Forskning från North Carolina finner att skolor som endast har motsvarande låg- och mellanstadium har ett något högre förädlingsvärde bland elever i sjätte klass, än skolor som även har högstadium. Det finns även stöd för att effekterna framför allt drivs av att lärare jobbar hårdare i skolorna med låg- och mellanstadium, även om en viss del även kan förklaras av att rektorer flyttar lärare till klassrum beroende på deras effektivitet (Macartney 2016).

Detta är dock framför allt ett problem i system där starka incitament knyts till skolors förädlingsvärden över specifika årskurser, vilket har varit fallet i North Carolina eftersom lärare där får bonusar om skolor når specifika mål. Men om de offerar förädlingsvärden i lägre årskurser för förädlingsvärden i högre årskurser syns det naturligtvis också i statistiken, vilket i sin tur minskar incitamenten att göra detta. Problemet skulle troligtvis ganska enkelt kunna åtgärdas genom att basera bonusarna på ett visst förädlingsvärde i alla årskurser – eller att inte använda bonusar över huvud taget. Som alltid påverkas incitamenten i ett system som betonar förädlingsvärden av hur myndigheter väljer att använda det (se OECD 2008), vilket delvis bör anpassas efter hur förädlingsvärdena har utformats.

Med andra ord är inte system som baseras på förädlingsvärden helt felfria, men de är svårare att manipulera än system som baseras på absoluta provresultat. Att informationssystem som baseras på förädlingsvärden är att föredra visas också i en annan studie av Macartney m.fl. (2015), som jämför effekterna av incitament som knyts till skolors förädlingsvärden med de som knyts till andelen som når godkänt. Resultaten visar att incitament som knyts till förädlingsvärden förbättrar de genomsnittliga resultaten 75 procent mer än incitament som knyts till andelen elever som når godkänt.⁶

Frågan är dock inte endast huruvida lågpresterande skolor höjer sig på grund av de incitament som skapas av informationssystem som baseras på förädlingsvärden. En annan viktig fråga är vad som händer när man använder sådan information för att ta beslut kring huruvida skolor måste stängas. Att stänga skolor innebär ofta att elever måste flytta till andra skolor, vilket kan inverka negativt på elevers resultat. Samtidigt bör detta ställas mot att dessa elever får möjlighet att gå på högre presterande skolor samt att framtida elever inte kommer utsättas för de sämre skolorna (se Heller-Sahlgren och Jordahl 2016). Forskning från Ohio som specifikt analyserar vad som händer när man stänger skolor med låga förädlingsvärden – vilket gjorde att eleverna där hamnade på skolor med högre förädlingsvärden – finner starka positiva effekter bland de elever som förflyttades från skolorna med låga förädlingsvärden (Carlson och Lavertu 2016). Den totala effekten av att stänga skolor med låga förädlingsvärden överväger alltså enligt denna forskning eventuella störningselement.

Likaså finner studier att incitamentssystem som fokuserar på förädlingsvärden också har möjlighet att höja lärarkårens totala effektivitet. I Washington DC finner exempelvis forskning att ett incitamentssystem som i hög grad baseras på lärares förädlingsvärden ledde till att lärare med låga förädlingsvärden försvann och ersattes av lärare med högre förädlingsvärden i genomsnitt. Detta ledde i sin tur till ganska stora positiva effekter på elevers provresultat (Adnot m.fl. 2017). Genom att offentliggöra lärares förädlingsvärden i Los Angeles tycks man också ha ökat sannolikheten att mer effektiva lärare stannar kvar i skoldistriktet jämfört med andra lärare. Samtidigt förbättrades också elevers resultat bland lärare med lägst förädlingsvärden, vilket tyder på att dessa lärare förbättrade sig efter publiceringen (Pope 2015).

⁶ Samtidigt ökar förädlingsvärdet spridningen i resultat. Detta på grund av att incitamenten som knyts till godkänt-gränsen ger starkare incitament att fokusera på lågpresterande elever, vilket minskar spridningen men heller inte förbättrar resultaten bland högre presterande elever i samma utsträckning.

Totalt sett finns det alltså mycket som tyder på att skolsystemet fungerar bättre med verktyg som baseras på incitament som knyts till skolors och lärares förädlingsvärden. En viktig fråga är också huruvida förädlingsvärdet påverkar föräldrars och elevers skolval. Överlag finns det en hel del som tyder på det. Ett sätt att utvärdera detta är att analysera hur huspriserna påverkas av skolors förädlingsvärden. Flera studier indikerar att föräldrar värderar skolors förädlingsvärde (och elevsammansättningen) när de köper hus (Brasington och Haurin 2006; Downs och Zabel 2002; Gibbons och Silva 2011; Gibbons m.fl. 2013; Martinez 2010).⁷ Viss amerikansk forskning som studerar föräldrars skolval finner också att skolors effektivitet spelar roll för vilka skolor de väljer, även om detta inte är den huvudsakliga slutsatsen som dras (Abdulkadiroglu m.fl. 2017).⁸ Andra studier finner dock mer blandade effekter (till exempel Cullen m.fl. 2006; Deming m.fl. 2014). I England är föräldrar samtidigt mer nöjda med skolor som har höga förädlingsvärden akademiskt sett, än de som är bäst på att öka deras elevers nöjdhet (Gibbons och Silva 2011).⁹ Forskningen som inte studerar förädlingsvärden men som justerar för elevernas bakgrund tyder samtidigt ofta på att föräldrar värderar sådana mått i sina skolval (se Heller-Sahlgren och Jordahl 2016). Samtidigt är det oklart hur informerade föräldrar är om skolors effektivitet i dessa studier.

Vad händer då när föräldrar får tillgång till mer information om skolors effektivitet? I ett intressant svenskt experiment i Linköping finner forskare att föräldrar börjar välja högstadieskolor med högre meritvärden, som justeras för elevernas bakgrund, när de får brev hemskickade med denna information. Intressant nog påverkas föräldrar inte av ojusterade resultat, vilket tyder på att de föredrar mer effektiva skolor snarare än skolor som har en mer fördelaktig elevsammansättning (Kessel och Olme 2018). Som diskuteras i avsnitt 5 finns det problem med att mäta förädlingsvärden genom att justera för elevers bakgrund på skolnivå i stället för deras initiala resultat på elevnivå, men analysen visar i varje fall att svenska föräldrar påverkas av ny information som man hoppas åtminstone delvis fångar upp skolors effektivitet.

I New York har man också funnit att publiceringen av enskilda lärares förädlingsvärden påverkar huspriserna i förväntad riktning: högre förädlingsvärden höjer huspriserna (Rodas 2015). Intressant nog finner viss forskning att publiceringen av lärares förädlingsvärden i Los Angeles i stället påverkade sorteringen av elever inom skolorna; högpresterande elever hamnade i högre utsträckning hos lärare med höga förädlingsvärden som publicerades, än hos lärare med samma förädlingsvärden som inte publicerades.¹⁰ Denna effekt är starkare när variationen i lärarkvalitet är hög inom skolan, vilket tyder på att föräldrarna är mer måna om att eleverna får en bra lärare när skillnaderna i lärarkvalitet inom skolan är stor och de inte har möjlighet

⁷ Brasington och Haurin (2006) samt Downs och Zabel (2002) påstår att de inte finner stöd för att förädlingsvärdet spelar, men detta baseras helt enkelt på en feltolkning av resultaten (se Heller-Sahlgren och Jordahl 2016).

⁸ Abdulkadiroglu m.fl. (2017) finner att föräldrar/elevers val i gymnasieansökningar påverkas av skolors effektivitet – som i sin tur påverkar vilka slags elever som söker till de olika skolorna. När författarna kontrollerar för kamraternas initiala resultat försvinner effekten av skolkvalitet, men detta förfarande riskerar helt enkelt att justera bort effekten av intresse (se Heller-Sahlgren 2017). Deras huvudsakliga resultat är att föräldrar värderar skolor som förbättrar deras barns resultat.

⁹ I Amsterdam tycks föräldrar välja skola efter hur höga resultaten är när de väljer högstadieskolor, även när man tar hänsyn till provresultaten i motsvarande mellanstadiet bland eleverna som togs in (Ruijs och Oosterbeek 2017). Detta tyder också på att föräldrarna värderar skolor med högre förädlingsvärden, även om analysen jämför initiala resultat bland andra elever än de som skolornas resultat beräknas efter.

¹⁰ Imberman och Lovenheims (2016) analyserar förändringar i huspriser i Los Angeles efter publiceringen av lärares och skolors förädlingsvärden och finner inga effekter. Det är dock omöjligt att veta om det är märkliga metodologiska val som driver resultaten – och det går även att tolka dem som att förädlingsvärdet faktiskt spelar roll (se Heller-Sahlgren och Jordahl 2016).

att byta skola (Bergman och Hill 2018).¹¹ Annan forskning finner dock inga genomsnittliga effekter på sorteringen av olika elever till lärare med olika kvalitet efter publiceringen (Pope 2015), vilket tyder på att likvärdigheten i tillgång till effektiva lärare inte påverkades märkbart överlag. Forskningen som studerar effekterna av att offentliggöra information om skolors resultat som justeras efter elevers bakgrund indikerar också att föräldrar ofta börjar välja bättre skolor efteråt (Heller-Sahlgren och Jordahl 2016).

Intressant nog finns det också stöd för att rektorer som ges information om lärares förädlingsvärden uppdaterar sina subjektiva omdömen om samma lärares effektivitet. Forskning från New York visar att rektorer som får information om lärarnas förädlingsvärden beräknade efter objektiva data påverkade deras subjektiva omdömen om lärarnas effektivitet. Detta gjorde i sin tur att mindre effektiva lärare i högre utsträckning förlorade jobbet, vilket hade positiva effekter på elevernas provresultat året efter (Rockoff et al. 2012). En annan studie från Houston, Texas, finner också att sannolikheten att mindre effektiva lärare förlorade arbetet ökade efter att man införde ett system som gav rektorer information om lärares förädlingsvärden (Cullen et al. 2017). Detta visar att rigorösa mått på förädlingsvärden kan vara av stor nytta för ledarskapet i skolan. Detta är dock inte speciellt förvånande eftersom det har visat sig vara väldigt svårt att kunna identifiera lärares effektivitet med blotta ögat, vilket noterades i avsnitt 2.

Det finns alltså starka tecken på att informationssystem som baseras på förädlingsvärden är att föredra framför system som fokuserar på absoluta resultat, både för att det minskar möjligheterna bland skolor och lärare att manipulera resultaten och för att det helt enkelt skapar starkare incitament att förbättra resultaten bland alla slags elever. Likaså verkar det som att aktörer inom skolsystemet uppdaterar sina subjektiva omdömen om lärares effektivitet när de får tillgång till objektiv information som baseras på förädlingsvärden. Tillgång till skolors och lärares förädlingsvärden är helt enkelt avgörande för att ens kunna veta vilka skolor och lärare som håller hög kvalitet och vilka som håller låg kvalitet. Som beskrivs i nästa avsnitt har man i Sverige tyvärr inte varit speciellt intresserad av att säkerställa att aktörerna inom skolsystemet har tillgång till sådana mått på skolors och lärares kvalitet.

¹¹ Detta stämmer överens med Jacob och Lefgrens (2007) resultat, som tyder på att föräldrar värderar lärare som förbättrar elevers resultat i skolor med sämre inputs, medan föräldrar som har att välja mellan skolor med bättre inputs inte värderar effektiva lärare i samma utsträckning.

5. Svenska mått på utbildningskvalitet och deras brister

Rapportens tidigare avsnitt har diskuterat vikten av att mäta skolors och lärares förädlingsvärden. Det är därför viktigt att notera att det i dag inte finns några mått på skolors eller lärares förädlingsvärden på nationell nivå. Visst publicerar Skolverket (2018a) information om skolors genomsnittliga resultat på sin hemsida, både för grund- och gymnasieskolor. Denna information syftar framför allt till att tillhandahålla underlag för analyser och jämförelser. För att underlätta föräldrars och elevers skolval har Skolverket (2018b) även utvecklat den nyare webbplatsen ”Välja skola”, som presenterar olika mått för skolors resultat – bland annat elevernas genomsnittliga meritvärde, andelen elever med godkänt betyg i alla ämnen, rapporter från Skolinspektionen samt resultatet från nöjdhetenkäter bland föräldrar, elever och lärare. Dessa mått lider dock av de problem som vi har diskuterat tidigare i rapporten: de fångar inte nödvändigtvis upp skolors effektivitet, utan snarare hur högpresterande deras elever är och vilken bakgrund de har.¹²

Skolverket har ett enda analysverktyg – Salsa – som justerar skolors genomsnittliga resultat i årskurs 9 för elevernas bakgrund, mätt som andelen flickor, andelen nyanlända elever och föräldrarnas genomsnittliga utbildningsbakgrund. Utfallsmåtten som används är andelen elever som når målen och elevernas genomsnittliga meritvärden. Liknande mått existerar inte i tidigare skolår eller på gymnasiet. Som diskuterades i avsnitt 4 finner forskning att föräldrar i Linköping började välja bättre högstadieskolor när de fick information om skolors Salsa-värden hemskickad (Kessel och Olme 2018). Detta tyder på att föräldrar värderar informationen men att de nödvändigtvis inte vet att den finns tillgänglig, eller att de inte vill lägga ner tiden som krävs för att ta reda på den. Men det är kanske inte så konstigt då Skolverket (2018c) menar att man inte bör använda Salsa för att rangordna skolor, då man påstår att kvalitet inte kan mätas i ett enda mått. Detta förklarar sannolikt varför man har valt att inte inkludera Salsavärdena bland de huvudsakliga kvalitetsmått som finns tillgängliga på den hemsida man har skapat för att underlätta föräldrars skolval.

Det är därför inte heller konstigt att Skolinspektionen inte tycks lägga större vikt vid huruvida skolorna de inspekterar är effektiva. Som beskrevs i avsnitt 2 har det visat sig svårt att genom observationer mäta skolors och lärares effektivitet när det gäller inläringen, men Skolinspektionen försöker inte ens utvärdera skolors effektivitet i sina observationer. Skolinspektionens tillsynsbeslut grundas i stället huvudsakligen på huruvida skolor följer skollagen, läroplaner och andra dokument gällande statens krav på skolsystemet; hur bra skolorna är på att förbättra elevernas kunskaper ligger inte i fokus. Det är symptomatiskt att man inte ens tar hänsyn till Salsa-värden när man i rapporterna diskuterar kunskapsmässig måluppfyllelse, utan i stället helt fokuserar på ojusterade resultat som inte tar någon hänsyn alls till skillnader i elevunderlagen. På gymnasienivå tar man inte heller hänsyn till elevernas initiala meritvärden som de har med sig från årskurs 9, vilket vore rimligt om man är intresserad av att fånga upp hur effektiva gymnasieskolorna faktiskt är. Givet att skolor inte kan göra någonting åt elevernas bakgrund är detta relativt uppseendeväckande.

¹² Sannolikt fångar enkäter upp viktiga icke-kognitiva färdigheter. Problemet med selektion är dock lika viktigt här. Vissa elever kommer exempelvis att vara nöjda oavsett vilken skola de går på, vilket gör det svårt att tolka resultaten i enkäterna som skolornas kausala effekter.

Det är samtidigt viktigt att notera att det även finns problem med Salsa, som alltså är det enda måttet med vilket man försöker fånga svenska grundskolors effektivitet. För det första inkluderar Salsa inget mått på elevernas tidigare resultat, vilket forskningen som diskuteras i avsnitt 3 visar är viktigt för att skapa tillförlitliga mått. För det andra skapas måttet från data på skolnivå, vilket gör att man riskerar att justera för en del av den kvalitet som man försöker fånga upp (se Heller-Sahlgren 2017). Dessutom tar måttet ingen hänsyn till att skolor erbjuder olika årskurser i grundskolan. Det gör att vissa elever vars resultat Salsa baseras på har varit på sina skolor sedan första klass, medan andra endast har varit där sedan sjunde klass. Salsa-liknande mått är sannolikt bättre än ojusterade resultat som informationsmått, men det är långt ifrån säkert att de på ett tillfredsställande sätt replikerar skolors förädlingsvärden.

Det är också värt att notera att alla kvalitetsmått som finns endast existerar på skolnivå. Man har struntat i att försöka mäta lärares effektivitet. Rektorer och myndigheter kan alltså inte jämföra olika lärares effektivitet, varken bland de som söker jobb på skolorna eller bland de som redan arbetar där. Dessutom finns den största variationen i lärares förädlingsvärden inom skolor snarare än mellan dem (till exempel Chetty m.fl. 2014a, 2014b), vilket indikerar att mått på enskilda lärares effektivitet skulle vara värdefulla för att få en bättre uppfattning om varför olika skolor presterar som de gör.

Ovanstående gäller mått som finns på nationell nivå. Det finns dock vissa kommuner som har försökt gå något längre i utformningen av mått som fångar upp skolors förädlingsvärden. Ett exempel är Stockholms stads (2018) utbildningsförvaltning, som har tagit fram ytterligare mått bland skolorna i kommunen. Dessa används endast för analys och har ännu inte presenterats offentligt. Där har man tagit fram ett Salsa-liknande mått som kallas ”Stockholms salsa”. Detta mått justerar elevers betyg i årskurs 9 med ett socioekonomiskt index som används för viktningen av skolpengen.¹³ Man justerar samtidigt för andelen elever med olika nivåer på detta index på skolan samt i det geografiska området eleverna bor i. I de jämförelser som har gjorts konstaterar man att det inte finns några större skillnader mellan hur skolor presterar enligt Salsa och enligt ”Stockholms salsa”, även om det är oklart hur man har kommit fram till denna slutsats. Måttet inkluderar fler bakgrundsvariabler än Salsa, men lider också av liknande problem som diskuteras ovan.

Sedan 2017 har Stockholms stad (2018) också skapat ett eget mått på högstadieskolors förädlingsvärde, som baseras på skolornas genomsnittliga meritvärde i årskurs 7. Detta jämförs helt enkelt med det genomsnittliga meritvärdet i årskurs 9. Endast de elever som har betyg från samma skola under årskurs 7, 8 och 9 ingår. Hittills har man endast räknat fram detta mått för en årskull elever. Problemet med detta mått är dock att eleverna redan har gått på högstadiet under ett år, vilket gör att man inte kan använda det för att fånga upp förädlingsvärdet i årskurs 7. Man beräknar också måttet på skolnivå i stället för på elevnivå, vilket som beskrivs ovan inte är perfekt. Dessutom bör man justera för betygen i årskurs 7 direkt i analysen för att ta hänsyn till de initiala resultatens betydelse för betygsutvecklingen – i stället för att endast mäta den absoluta förändringen över tid – enligt forskningen som diskuteras i avsnitt 3.

Till sist mäter Stockholms stad (2018) förädlingsvärden bland gymnasieskolornas nationella program. I modellen används elevernas meritvärden från årskurs 9 för att uppskatta hur många elever som bör nå gymnasieexamen bland eleverna som finns

¹³ Detta index bygger på en indikator för om eleven är nyinvandrad, graderad efter hur lång tid eleven har varit i Sverige (upp till sju år, vårdnadshavarnas utbildningsbakgrund, om vårdnadshavarna uppstår ekonomiskt bistånd och huruvida eleven bor med båda, en eller inga vårdnadshavare.

på skolan i årskurs 3. Man har beräknat sannolikheten att en elev med ett specifikt meritvärde tar gymnasieexamen. Man har inte genomfört samma analys med den genomsnittliga betygspoängen, men noterar att detta också kan vara av intresse.¹⁴

Stockholms stad (2018) tar dock samtidigt aktivt avstånd från att skapa förädlingsvärden efter resultaten på de nationella proven – både i grundskolan och på gymnasienivå – eftersom dessa endast fångar upp vissa kunskapsområden och eftersom innehållet kan variera mellan olika år. Detta är en märklig bedömning eftersom proven skapas efter analyser av läroplanen samt ämnes- och kursplaner och är utformade för att täcka så stora delar av kursplanen som möjligt. Som noteras i avsnitt 8.4 kan vissa variationer i proven över tid dessutom vara en fördel, eftersom man då bättre kan säkerställa att lärare och skolor inte endast coachar elever för att klara av ett specifikt slags prov.¹⁵ På grund av lärarnas autonomi är variationen i både innehållet och kunskapskontrollerna som betygen grundas på också mycket större än eventuella (och slumpmässiga) variationer i de centralt utformade nationella proven – som trots allt är de enda enhetliga kunskapskontrollerna som existerar i den svenska skolan.

Hur som helst har Stockholms stad alltså tagit fram mått som ämnar fånga upp skolors effektivitet vad gäller betyg, vilka ännu inte existerar på nationell nivå. Fast det naturligtvis finns problem med måtten är de bättre än ingenting. Samtidigt är det värt att betona att dessa indikatorer för tillfället endast används för analys; man anser att osäkerheten är för stor – eftersom måtten ”är metodkänsliga och kan vara svårtolkade” – för att man ska offentliggöra de mått på skolors effektivitet som man har tagit fram. Man är också rädd för att de ska öka betygsinflationen. Detta trots att man samtidigt menar att måtten ”på ett godtagbart sätt belyser skolornas bidrag till elevernas kunskapsutveckling” (Stockholms stad 2018, s. 2).

Totalt sett står det alltså klart att det råder informationsbrist kring skolors och lärares effektivitet i Sverige. Det är därför inte speciellt konstigt att svenska föräldrar upplever att deras skolval försvåras av bristande information (Malmberg m.fl. 2013). Historiskt sett har det varit svårt att skapa mått på skolors förädlingsvärden i grundskolesystemet på nationell nivå, eftersom man länge samlade in data på elevers resultat för första gången i årskurs 9. Men man har inte försökt skapa bättre mått på gymnasienivå heller, trots att man där har haft möjligheter att ta hänsyn till elevernas initiala resultat. Myndigheter använder generellt inte heller den information om skolors faktiska effektivitet som de själva har skapat (Salsa). Allt sammantaget tyder detta på en generell ovilja att ta tag i frågan på allvar.

Denna ovilja blir än tydligare när man tar i beaktning att Skolverket på uppdrag av Utbildningsdepartementet (2012) faktiskt utredde möjligheten att mäta skolors förädlingsvärden – men valde att inte gå vidare med idén, eftersom forskarna man konsulterade och Skolverkets egna experter varnade för eventuella problem. Man ansåg helt enkelt att det inte fanns utfallsmått som var lämpliga för att skapa sådana mått. Man var speciellt rädd för att måtten skulle leda till betygsinflation. Man var också rädd för att små skolenheter skulle skapa osäkerheter i de beräknade värdena (Arevik 2014; Hagström 2015; Skolverket 2015). I utlåtandet reflekterade man intressant nog dock endast över förädlingsvärdets betydelse vid skolval, inte hur ett sådant mått skulle kunna användas för skolutveckling mer generellt (Skolverket 2015). I sitt yttrande

¹⁴ Skolkedjan Academedia använder samma modell för att beräkna den förväntade genomsnittliga betygspoängen bland skolorna i kedjan.

¹⁵ Stockholms stad (2018) menar att skolor kan inrikta undervisningen på olika områden vid olika tidpunkter, vilket kan slå igenom på resultaten. Men detta gäller alla prov, oavsett utformning, och sådana effekter kan också ses som en mekanism genom vilken skolorna bidrar till resultaten; att rikta undervisningen på ett visst sätt kan vara mer effektivt än att rikta undervisningen på ett annat sätt.

över Skolkommissionens rapport – där man ”särskilt betona[ade] betydelsen av att det utvecklas mått för en skolas bidrag till elevernas kunskapsutveckling” (SOU 2017:35, s. 295) för att underlätta skolvalet – menade Skolverket (2017) att man vidhåller sin tidigare bedömning att det är olämpligt att bygga sådana mått på betyg eller resultat från nationella prov.

Som avsnitt 8.1 diskuterar är dagens interna betygssättning mycket riktigt ett av flera hinder som gör det svårt att skapa rigorösa mått på skolors effektivitet, men det är svårt att se att incitamenten till inflation i sig skulle vara starkare om man tar fram mått på skolors förädlingsvärden. Redan i dag bedöms skolor efter hur de presterar på nationella prov samt vilka betyg de når i genomsnitt, vilket framkommer tydligt i Skolinspektionens rapporter. Det är också dessa resultat som ligger i fokus som kvalitetsmått på Skolverkets egen hemsida som ämnar underlätta föräldrars skolval. Med ett större fokus på förädlingsvärden skulle dessa incitament inte försvinna, men man skulle få en större förståelse för att skolors förutsättningar skiljer sig åt på grund av att de har elever med olika förutsättningar. Att helt strunta i möjligheterna att utforma mått på skolors förädlingsvärden inom ramen för nuvarande system, med hänvisning till incitamenten till betygsinflation, ter sig därför väldigt märkligt. Detsamma gäller osäkerheten som kan uppstå på grund av små skolenheter. Frågan är helt enkelt inte om måttet skulle vara perfekt, utan huruvida det är bättre att inte ha något sådant mått över huvud taget och i stället förlita sig på skolors ojusterade resultat.

Naturligtvis är det viktigt att notera begränsningarna som nuvarande system sätter för möjligheterna att utforma ett bra mått på skolors och lärares förädlingsvärden. Dessa begränsningar diskuteras mer i detalj i avsnitt 8. Dagens system är inte utformat för att man med stor säkerhet ska kunna mäta skolors och lärares förädlingsvärden. Men samtidigt står det också klart att man faktiskt kan skapa bättre mått än vad som finns tillgängliga i dag. Detta diskuteras mer i detalj i avsnitt 7. Men innan dess är det värt att exemplifiera hur andra skolsystem har skapat och offentliggjort förädlingsvärden som mått på skolors och lärares effektivitet.

6. Andra länders mått på förädlingsvärden

Det här avsnittet diskuterar hur man i tre andra länder använder förädlingsvärden i skolan. Detta för att exemplifiera hur andra länder har skapat sådana mått och vad Sverige kan lära av dessa länder.

6.1 Norge

Historiskt sett har man inte mätt skolors förädlingsvärden i Norge, men detta har nyligen förändrats. De nya måtten uppstod först i forskning som analyserade hur man skulle kunna utforma skolors förädlingsvärden på gymnasienivå, bland annat av Falch m.fl. (2016) och Markussen m.fl. (2017). Forskningen ledde fram till att det norska skolverket (Utdanningsdirektoratet) i december 2018 för första gången publicerade mått på gymnasieskolors förädlingsvärden på sin hemsida vad gäller ökningen i genomsnittliga resultat, andelen elever med godkända betyg i alla ämnen under året som värdet gäller och andelen som inte hoppar av under året som värdet gäller. Måtten beräknas separat för elever i årskurs 1, 2 och 3 på gymnasiet, och man justerar alltså för elevernas resultat under tidigare år. Dessa mått presenteras årligen i förhållande till genomsnittet i landet, med konfidensintervall för att ta hänsyn till osäkerheten i resultaten. Måtten publiceras separat för studieförberedande program och yrkesprogram och justeras för elevernas kön, vilket program eleverna läser på och vilken årskurs de befinner sig i. Likaså presenteras måtten aggregerade på kommunnivå. Förädlingsvärdena baseras på utfall som bestäms av elevernas egna lärare, men man har justerat för eventuell betygsinflation genom att ta hänsyn till skillnaden mellan betygen och resultaten i de centralt rättade proven som skrivs i grundskolan och gymnasiet (se Utdanningsdirektoratet 2018).

Hittills har Utdanningsdirektoratet inte skapat liknande mått på grundskolenivå. Men år 2017 valde den norska statistiska centralbyrån att analysera grundskolors och kommuners effektivitet och skapa förädlingsvärden som baseras på de nationella proven i årskurs 5, 8 och examensprovet i årskurs 10 i engelska, norska och matematik (Steffensen m.fl. 2017). Det genomsnittliga resultatet i alla tre ämnen används för att ta fram måttet. Proven som används har naturligtvis vissa brister; exempelvis är de inte utformade för att fånga upp alla kunskaper som man vill inskräpa i de olika ämnena. Men de nationella proven rättas digitalt och automatiskt, förutom ett par uppgifter i läsning och examensproven i årskurs 10 som rättas externt. För att minska eventuellt mätfel slår man samman två års resultat och utesluter skolor med färre än 20 elever totalt sett över de två åren. Man använder både tidigare resultat och elevernas bakgrund som kontroller i modellen, men i enlighet med studierna som diskuteras i avsnitt 3 är resultaten i princip likadana om man exkluderar bakgrundsfaktorerna.

Forskarna presenterar förädlingsvärden för landets alla skolor och kommuner mellan årskurs 5 och 8 samt mellan årskurs 8 och 10. De diskuterar för- och nackdelar med att göra detta, men landar i att fördelarna väger tyngre än nackdelarna. Alla resultat presenteras också med felmarginaler, så att man kan analysera skillnaderna mellan olika skolor med 95 procents säkerhet. De flesta skolor skiljer sig inte från det nationella genomsnittet. Detta är dock inte det mest relevanta i sammanhanget, eftersom

man väljer mellan olika skolor varav de flesta naturligtvis inte presterar exakt i nivå med snittet. Det är inte heller avgörande att skolor skiljer sig åt med 95 procents säkerhet; så länge de skiljer sig med mer än 50 procents säkerhet är det bättre att använda måttet än att inte använda sig av det.

Överlag står det alltså klart att Norge har kommit mycket längre än Sverige när det gäller att mäta skolors förädlingsvärden. Utdanningsdirektoratet har gjort en annan bedömning än Skolverket vad gäller fördelarna med att skapa mått på skolors förädlingsvärden. På gymnasienivå finns det nu tre mått som på olika sätt fångar upp skolors förädlingsvärden under gymnasiets alla tre årskurser, vilket får anses vara ambitiöst. Utdanningsdirektoratet har ännu inte börjat framställa förädlingsvärden för grundskolan, men sådana värden har redan tagits fram och offentliggjorts av forskare, som beskrevs ovan. Det är därför relativt enkelt för Utdanningsdirektoratet att publicera liknande mått i framtiden om man så vill. Liksom i Sverige finns det dock inga möjligheter att skapa förädlingsvärden för enskilda lärare, eftersom det inte via centrala register går att koppla en enskild lärare till eleverna de undervisar.

6.2 England

Informationsutbudet vad gäller skolors förädlingsvärden i England är idag förhållandevis gott. I början på 2000-talet infördes förädlingsvärden som konstruerades efter de prov som skrivs i årskurs 2, 6, 9, 11 och 13. Med hjälp av dessa prov skapade man förädlingsvärden över motsvarande låg- och mellanstadiet, högstadiet samt gymnasiet. Dessa mått justerades under ett par år (2006–2010) även för bakgrundsfaktorer på elevnivå och för skolornas genomsnittliga initiala resultat (se till exempel DCSF 2003a, 2003b, 2007). År 2011 slutade man dock justera för elevers bakgrund och skolornas genomsnittliga resultat eftersom den nya regeringen ansåg att måttet var svårt att förstå samt skapade låga förväntningar för elever från mindre bemedlade hem (DfE 2010). År 2010 avskaffades också proven i årskurs 9, vilket gör att man sedan dess endast beräknar måtten på förädlingsvärden i högstadiet mellan årskurs 6 och 11 (DCSF 2009). Skolors förädlingsvärden offentliggörs årligen av det engelska utbildningsdepartementet och publiceras även i nationella tidningar. Under de senaste åren har förädlingsvärdet mellan årskurs 2 och 6 samt mellan årskurs 6 och 11 också rapporterats uppdelat efter elevers tidigare resultat – och efter huruvida de är fattiga eller inte – vilket gör det möjligt att skilja på skolor som är olika effektiva för olika slags elever. Förädlingsvärdet mellan årskurs 11 och 13 har endast rapporterats uppdelat efter huruvida eleverna är fattiga eller inte.

Trots att man har publicerat förädlingsvärden under rätt lång tid tenderade man fortfarande länge att ge störst uppmärksamhet åt skolors genomsnittliga resultat. Detta har dock förändrats, i varje fall i högstadiet och gymnasiet. Nyligen gick man över till att betona ett nytt förädlingsvärde mellan årskurs 6 och 11 (Progress 8), som har skapats genom att man jämför elevers resultat i åtta ämnen i årskurs 11 med de genomsnittliga resultaten på nationell nivå bland elever som hade liknande resultat i proven i läsning, skrivning och matematik i årskurs 6. Resultaten i årskurs 11 utgör sedan en del av måttet på förädlingsvärde som används i motsvarigheten till gymnasiet, där den andra delen utgörs av elevernas resultat i slutproven i årskurs 13. Likaså beräknar man förädlingsvärden mellan årskurs 2 och 6 i läsning, skrivning och matematik. Bägge dessa mått skapas på ett liknande sätt som Progress 8. Alla mått beräknas på elevnivå och aggregeras sedan på skolnivå, med ett konfidensintervall inom vilket skolorna befinner sig med 95 procents säkerhet (se DfE 2018a, 2018b, 2018c).¹⁶

¹⁶ Se Leckie och Goldstein (2017) för en längre diskussion om hur måtten på förädlingsvärden har förändrats över tid i England.

Det är oklart i vilken utsträckning de äldre och nuvarande måtten på förädlingsvärden faktiskt fångar upp skolors effektivitet och i vilken utsträckning de kan förutspå långsiktiga utfall.¹⁷ En styrka i måtten är dock att man numera redovisar skolors förädlingsvärden uppdelat efter elevers initiala resultat, vilket tar hänsyn till att skolor kan vara olika effektiva för olika slags elever. Eftersom det inte via centrala register går att koppla lärare till eleverna de undervisar finns det samtidigt inga mått som mäter enskilda lärares förädlingsvärden.

Sammantaget finns det mycket som tyder på att de engelska måtten är ganska bra, men att man sannolikt skulle behöva förbättra dem i vissa avseenden. Exempelvis finns det en risk att avståndet mellan det första och andra mättillfället i grundskolan skapar onödigt hög instabilitet i måtten. Inte heller finns det möjligheter att mäta enskilda lärares förädlingsvärden i det engelska skolsystemet. Hur som helst framstår de engelska kvalitetsmåtten överlägsna de svenska eftersom man i Sverige inte tar hänsyn till elevernas initiala resultat över huvud taget.

6.3 USA

På grund av USA:s decentraliserade utbildningssystem finns det en mängd olika system för att säkerställa information om skolors och lärares effektivitet. Sedan den federala lagen ”No Child Left Behind” (NCLB) infördes 2002 måste alla skoldistrikt publicera skolors resultat i årliga rapporter som innehåller absoluta provresultat i olika former, jämförelser mellan skolor i distriktet och i USA generellt samt information om vilka skolor som är i behov av förbättring. Under senare delen på 2000-talet deltog en del delstater i ett pilotprogram där fokus istället låg på förädlingsvärde, och numera kan de också ansöka om att undantas från vissa krav i utbyte mot att lärare och rektorer utvärderas efter sina förädlingsvärden (Heller-Sahlgren och Jordahl 2016). Man har alltså börjat lägga större vikt vid förädlingsvärden på federal nivå än vad som har varit fallet historiskt sett.

Vissa delstater har dock fokuserat på skolors renodlade förädlingsvärden under ganska lång tid. År 1996 infördes exempelvis ett system för ansvarsutkrävande i North Carolina som fokuserade på 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 i skolan 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.¹⁸

¹⁷ En väldigt enkel analys indikerar att måtten på motsvarande högstadieskolors förädlingsvärden – som förutom elevernas initiala resultat också justerades för deras bakgrundsvARIABLER och skolors genomsnittliga initiala resultat – som publicerades under det tidiga 2000-talet tycks ha haft ett hyfsat prediktivt värde vad gäller långsiktiga utbildningsresultat men ganska lågt prediktivt värde vad gäller arbetsmarknadsutfall (Thomson 2018).

¹⁸ I vissa fall har dock privata aktörer valt att offentliggöra lärares förädlingsvärden. I Los Angeles publicerades exempelvis lärares och skolors förädlingsvärden, efter att *Los Angeles Times* anlidade forskaren Richard Buddin (2010) för att skapa förädlingsvärden baserade på provresultat i matematik och läsförståelse bland elever i årskurs 3–5 för alla lärare och skolor i skoldistriktet. Ett år efteråt valde skoldistriktet LAUSD sedan att beräkna och offentliggöra skolors förädlingsvärden, medan man endast gjorde lärarnas förädlingsvärden tillgängliga för lärare och rektorer (se Pope 2015). Även i New York City har enskilda lärares förädlingsvärden offentliggjorts. Utbildningsdepartementet hade länge beräknat sådana värden i matematik och engelska för årskurserna 4–8, vilka användes internt för att bestämma vilka lärare som blev befordrade. I februari 2012 publicerades dock resultaten för åren 2007–2010 av New York Times. Sedan dess har systemet dock förändrats för hela delstaten New York och baseras i dag på både förädlingsvärden (i fler ämnen) och observationer av rektorer, men lärare som klassas som ineffektiva enligt förädlingsvärdet kan inte få ett högre betyg i genomsnitt (se Rodan 2015).

Generellt sett använder många delstater och skoldistrikt i dag skolors och lärares förädlingsvärden som en av flera komponenter inom ramen för officiella utvärderings-system. Kanske det mest iögonfallande exemplet är Washington DC, där man år 2009 införde ett extremt hårt system för ansvarsutkrävande. Inom ramen för systemet utvärderas alla lärare årligen, bland annat efter deras förädlingsvärden. Varje lärare får ett betyg och de som bedöms vara ”väldigt effektiva” får monetära och icke-monetära belöningar, medan de som bedöms vara ”ineffektiva” eller under två år i rad ”minimalt effektiva” sparkas. Systemet är väldigt kontroversiellt men har visat sig förbättra resultaten och lärarnas kvalitet markant (Adnot m.fl. 2017; Dee och Wykoff 2015).

Oavsett vad man tycker om systemen för ansvarsutkrävande kan man säga att många amerikanska delstater och skoldistrikt i vissa bemärkelser har kommit långt när det gäller utformningen av förädlingsvärden i skolsystemet, eftersom man även mäter sådana värden på lärarnivå. Detta gör att måtten i högre utsträckning kan användas av rektorer vid anställningar och utvärderingar av lärare på skolorna. Det finns samtidigt skillnader mellan hur delstater och skoldistrikt exakt använder förädlingsvärdena, och det är oklart vilket sätt som är mest fördelaktigt.

7. Hur kan man mäta förädlingsvärden i Sveriges skolsystem?

Hur kan man då skapa förädlingsvärden på nationell nivå i Sverige? Det här avsnittet diskuterar vad man kan göra inom ramen för dagens system och de data som finns att tillgå, och exemplifierar sedan med data från en svensk kommun. Nästa avsnitt diskuterar vad som krävs för att förbättra möjligheterna att mäta skolors och lärares förädlingsvärden i det svenska skolsystemet.

Som noterades i avsnitt 5 publiceras inga mått på skolors förädlingsvärden på nationell nivå i Sverige. Man har tidigare kunnat estimeras gymnasieskolors förädlingsvärden på olika program, genom att analysera elevers grundskolebetyg och deras slutbetyg i gymnasieskolan. Som beskrevs i avsnitt 5 görs detta i viss utsträckning i Stockholms stad, även om denna information inte offentliggörs. Historiskt sett har resultaten från nationella prov inte samlats in systematiskt på gymnasienivå, vilket har gjort det svårt att studera förädlingsvärden med dessa resultat. Nu kan man dock använda de nationella proven i årskurs 9 och nationella prov respektive betygen på gymnasiet för att estimeras olika skolors förädlingsvärde, totalt sett eller för olika program.

Efter alliansens reformer under åren 2006–2014 finns nu också större möjligheter att skapa mått på förädlingsvärden i grundskolan. Sedan läsåret 2008/09 skriver elever i årskurs 3 nationella prov i matematik och svenska, och sedan läsåret 2011/12 skriver elever i årskurs 6 nationella prov i engelska, matematik och svenska. I årskurs 3 samlar man in uppgifter om huruvida eleverna når kraven på de olika delproven, men även den exakta provpoängen. Detta gör att man kan beräkna skolors förädlingsvärden i matematik och svenska, men inte engelska, mellan årskurs 3 och 6. Eftersom elever inte skriver prov eller får betyg i förskolan – och inte heller skriver diagnostiska prov precis när de har börjat skolan – finns det dock inga möjligheter att beräkna förädlingsvärden i lågstadiet.

I årskurs 6 samlas resultaten in från alla tre nationella prov i engelska, matematik och svenska. Eftersom samma elever tre år senare skriver nationella prov i årskurs 9 finns därför nu möjligheten att utveckla ett mått på högstadieskolors förädlingsvärden i alla tre kärnämnen. Likaså kan man sedan betyg infördes i årskurs 6 år 2012/13 även beräkna skolors förädlingsvärden mellan årskurs 6 och 9 i den bemärkelsen. Fördelen med nationella prov är att alla elever utvärderas efter samma innehåll och kravnivåer, vilket inte nödvändigtvis stämmer när det gäller betyg givet lärarnas autonomi vad gäller innehåll och betyggrundande moment. I årskurs 6 betygssätts dessutom endast eleverna i de ämnen som lästs under just det året, vilket innebär att antalet ämnen som eleverna får betyg i varierar. Likaså är incitamenten att inflatera resultaten på nationella prov sannolikt mindre än på slutbetygen som eleverna får. Det finns dock naturligtvis fördelar med att använda betygen som grund för förädlingsvärdena: som noterades i avsnitt 3 tycks dessa mått fånga upp icke-kognitiva förädlingsvärden i högre utsträckning.

Naturligtvis bestäms varken betygen eller resultaten på de nationella proven externt, vilket skapar icke-likvärdig betygssättning (Skolinspektionen 2018). Detta är dock ett problem i hela det nuvarande betygs- och antagningssystemet. I avsnitt 8 diskuteras vad som bör göras för att man ska kunna skapa rigorösa mått mer generellt, medan jag här nöjer mig med att analysera vad som kan göras inom ramen för nuvarande system.

För att visa hur man kan göra detta i praktiken redovisas nedan resultaten från en analys över högstadieskolors förädlingsvärden på nationella prov i en svensk kommun. Även om resultaten endast gäller för högstadieskolor är de relevanta även för andra nivåer i utbildningssystemet, då man skulle kunna använda samma ansats för att ta fram förädlingsvärden mellan årskurs 3 och 6 samt mellan årskurs 9 och gymnasiet.

7.1 Förädlingsvärden från en svensk kommun

Analysen genomfördes av mig och professor Henrik Jordahl i ett försök att ta fram mer relevanta mått på skolors kvalitet i Sverige. Syftet med projektet var att studera och kartlägga högstadieskolors förädlingsvärden i kommunen som ett pilotprojekt; i framtiden kommer vi förhoppningsvis att studera alla skolor i Sverige.¹⁹ Här använder jag endast huvuddragen från analysen för att exemplifiera hur man kan beräkna skolors förädlingsvärden inom ramen för dagens system – samt i vilken utsträckning existerande kvalitetsmått tycks reflektera skolors effektivitet – och har därför valt att hålla både kommunen och skolorna anonyma. Analysen kommer att publiceras i sin helhet vid ett senare tillfälle.

För att beräkna skolors förädlingsvärden på ett korrekt sätt är det avgörande att använda data på elevnivå (se Heller-Sahlgren 2017). Förädlingsvärdena beräknas för alla skolor i kommunen som har elever i årskurs 9. Datamaterialet kommer från SCB och innehåller uppgifter för alla elever som var registrerade i årskurs 7, 8 eller 9 i en skola i kommunen under skolåren 2013/14–2016/17. Beräkningarna görs först på elevnivå och aggregeras sedan på skolnivå. Förädlingsvärdena redovisas alltså som genomsnitt för var och en av kommunens skolor över bägge åren.

Förädlingsvärdena baseras på elevernas provbetygspoäng i de nationella proven i svenska, engelska och matematik. De olika provbetygen ger poäng enligt följande: A ger 20 poäng, B ger 17,5 poäng, C ger 15 poäng, D ger 12,5 poäng, E ger 10 poäng och F ger 0 poäng.

Förädlingsvärdet baseras på skillnaden mellan provbetygspoängen i årskurs 9 och årskurs 6. Vi valde att basera förädlingsvärdena på resultaten i nationella prov snarare än betyg eftersom det sannolikt finns starkare incitament att inflatera betygen än de nationella proven och eftersom de nationella proven är den enda enhetliga kunskapskontrollen som existerar. Fördelen med nationella prov är att alla elever utvärderas efter samma innehåll och kravnivåer, vilket inte nödvändigtvis stämmer när det gäller betyg på grund av lärarnas autonomi vad gäller innehåll och betygsgrundande moment. Hur som helst kan man även skapa förädlingsvärden baserade på betygen i årskurs 6 och 9 med samma metod som redovisas här om man så vill.

Vi började med att skapa ett genomsnitt över resultaten på proven i de tre ämnena svenska, engelska och matematik för varje elev och årskurs. Som noterades i avsnitt 3 bör man också justera för elevens initiala resultat i måttet som elevens progression mäts utifrån för att fånga upp skolors och lärares bidrag till inläring, vilket är vad vi är intresserade av (se till exempel Chetty m.fl. 2014a; Deming 2014). Genom att justera för elevernas provbetyg i årskurs 6 tar vi hänsyn till elevernas tidigare resultat, som i sin tur påverkar skolornas möjligheter att bidra till inläringen ytterligare.²⁰

¹⁹ Projektet finansierades delvis av Watma Group AB (moderbolag till skolkedjorna Nordic International School och Fria Läroverken), som efterfrågade analysen som ett underlag för att stärka företagets kvalitetsarbete i skolsektorn generellt och för att analysera hur ett informationssystem baserat på skolors förädlingsvärde skulle kunna påverka incitamentsstrukturen på skolmarknaden.

²⁰ Vi experimenterade också med att lägga till de initiala resultaten i kvadrat och kubik, men valde att använda en linjär modell eftersom forskningen som diskuteras i avsnitt 3 generellt indikerar att detta är tillräckligt för att fånga upp skolors och lärares effektivitet.

Detta gör att estimaten per definition är relativa och att skolornas förädlingsvärden ska tolkas i relation till varandra. En skola vars elever presterar väl i årskurs 6 men försämrar sina resultat till årskurs 9 kan ändå ha ett positivt förädlingsvärde – om elever på andra skolor har försämrat sina resultat ännu mer. Genom att justera för provresultatet i årskurs 6 skapar vi en situation där alla skolor utgår från samma nivå. Och om eleverna på en skola då försämrar sina resultat från denna nivå mindre än eleverna på andra skolor försämrar sina resultat från samma nivå kommer modellen alltså ange att skolan har en positiv påverkan på elevernas resultat.

Enligt tidigare forskning är initiala provresultat tillräckliga för att skapa väntevärdesriktiga estimat av lärares och skolors bidrag till inläringen (Chetty m.fl. 2014a; Deming 2014). Men för att ytterligare säkerställa att vårt mått fångar upp skolors faktiska kvalitet analyserade vi i vilken utsträckning resultaten förändrades om vi utöver provresultatet i årskurs 6 även justerar för ett stort antal bakgrundsvariabler.²¹

Eleverna inkluderas oavsett var i Sverige de gick skolan i årskurs 6. De förädlingsvärden som vi redovisar avser skolor i kommunen och baseras på de elever som gick i årskurs 9 vid dessa skolor läsåren 2015/16 eller 2016/17. Dessa elever kan alltså ha gått årskurs 6 i en annan skola i kommunen eller i en skola belägen i en annan kommun.²²

7.1.1 Resultat

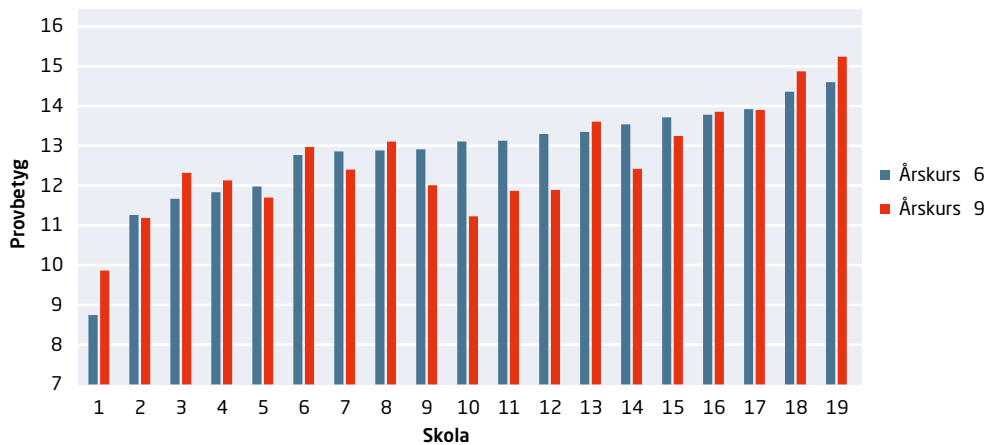
Figur 1 visar elevernas genomsnittliga provresultat i årskurs 6 och årskurs 9 på skolnivå. Det är tydligt att eleverna i de olika skolorna har olika utgångspunkter vad gäller resultat, och deras absoluta resultatförändringar varierar också kraftigt. Elever som förbättrar sig mest är de med lägst initiala provresultat. Elever i skola 1 presterar i genomsnitt under betyget E i de nationella proven i årskurs 6. Detta gör det till viss del lättare för eleverna att förbättra sina resultat, då de som höjer sig från F till E förbättrar sig med hela tio provpoäng. Samtidigt når eleverna i skola 1 i genomsnitt inte ens upp till gränsen för E på de nationella proven i årskurs 9. Trots att elever i denna skola förbättrar sina absoluta provresultat mest av alla, underpresterar de fortfarande i absolut bemärkelse.

Skolorna med de största absoluta resultatförsämringarna utgår däremot från mer genomsnittliga värden. Elevernas genomsnittliga provresultat i skola 10 och skola 11 försämrar exempelvis från ett snitt på över 13 provpoäng, vilket är strax över det kommunala medelvärdet. Samtidigt förbättrar eleverna i skola 18 och skola 19 sina resultat från initiala snitt på över 14 provpoäng, vilket är högst i kommunen. Trots dessa utgångslägen – från vilket förbättringar borde vara svårare än för skolor med sämre utgångslägen – är elevernas absoluta förbättringar i dessa skolor de fjärde och femte högsta i kommunen.

²¹ Dessa variabler är: kön, födelseland (regionindelad), eventuellt invandringsår, födelsekvartal, faderns födelseland (regionindelad), moderns födelseland (regionindelad), faderns disponibla inkomst, moderns disponibla inkomst, faderns utbildnings (7 kategorier) och moderns utbildning (7 kategorier). Bortfallet i kontrollvariablerna är generellt lågt, men för att säkerställa att vi använder så stor del av populationen som möjligt tillskriver vi observationer med bortfall på kontrollvariablerna värdet noll samt inkluderar indikatorer för bortfall.

²² I analysen som rapporteras här exkluderas en skola som hade färre än fem elever för vilka man kunde beräkna ett förädlingsvärde. Vi exkluderade denna skola i vår analys för att säkerställa att man inte kunde identifiera enskilda elevers resultat; på grund av riskerna för detta prickar Skolverket data som baseras på färre än 10 elever.

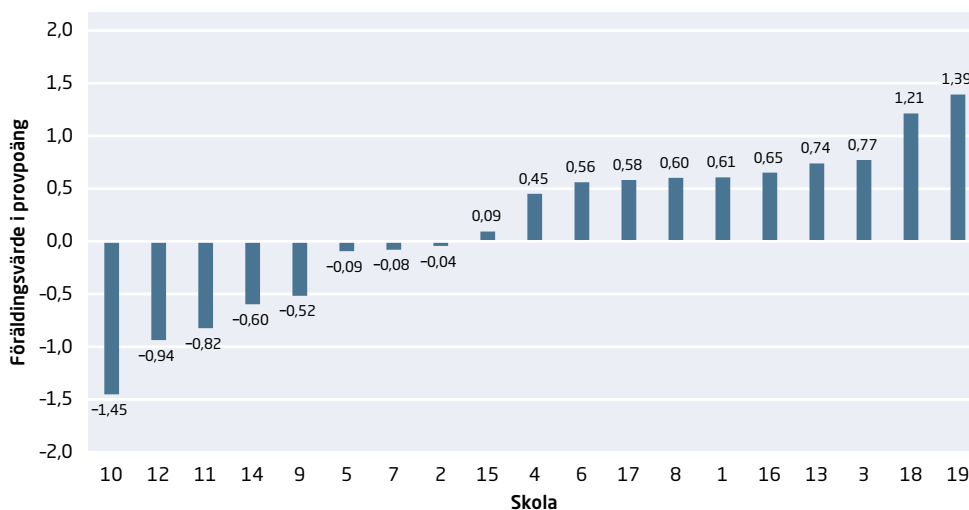
Figur 1. Genomsnittliga provbetyg i årskurs 6 och 9



Dessa skillnader reflekterar dock inte nödvändigtvis enbart skolornas bidrag till resultatutvecklingen. Som står klart är utgångsvärdet viktigt för progressionen. Figur 2 redovisar därför resultaten från modeller som justerar för elevernas genomsnittliga provbetyg i årskurs 6. Skolornas förädlingsvärde beräknas således av differensen mellan elevernas progression – skillnaden mellan provresultaten i årskurs 9 och årskurs 6 i figur 1 – och den som kan förutspås av elevernas provresultat i årskurs 6.

Resultaten visar att eleverna på skola 19 har det högsta förädlingsvärdet: i genomsnitt bidrar skolan med 1,39 provpoäng till elevernas resultat i årskurs 9. Detta motsvarar en ökning med 56 procent av ett betygssteg, räknat från E och uppåt, eller 14 procent av steget mellan F och E. Den lägsta progressionen återfinns i skola 10, som i genomsnitt reducerar elevernas resultat med 1,45 provpoäng relativt sett. Detta motsvarar en försämring med 58 procent av ett betygssteg räknat från E och uppåt, och 14 procent av steget från F till E. Skillnaden mellan skolorna med högst och lägst progression motsvarar därför 1,14 betygssteg räknat från E och uppåt, eller 0,28 betygssteg mellan F och E.

Figur 2. Skolornas förädlingsvärde i nationella prov mellan årskurs 6 och 9



Här rapporteras endast skolornas genomsnittliga förädlingsvärden i engelska, matematik och svenska. Men icke-rapporterade resultat tyder på att de relativa förädlingsvärdena varierar något beroende på när man analyserar de olika ämnena var för sig. Exempelvis når skola 4 ett förädlingsvärde på 1,15 i matematik, -0,23 i engelska och 0,11 i svenska.

Skillnaden mellan värdet i matematik och engelska motsvarar ett hopp i rangordningen från plats 5 till plats 13. Skola 11 återfinns på sista plats i engelska, med ett förädlingsvärde på $-1,22$, men uppnår plats 9 i matematik, med ett förädlingsvärde på $0,79$. Skola 19 presterar å andra sidan väldigt väl i alla tre ämnen. Men överlag indikerar mönstret alltså att skolors förädlingsvärden mycket väl kan variera beroende på ämne.

I analyser där skolors förädlingsvärden beräknades som differensen mellan den faktiska förändringen – skillnaden mellan provresultaten i årskurs 9 och årskurs 6 som redovisas i figur 1 – och den som kan förutspås av elevernas genomsnittliga provresultat i årskurs 6 samt bakgrundsfaktorerna, var resultaten överlag väldigt lika, även om ett par skolor hoppar något steg upp eller ner i rangordningen.

Självklart ska man komma ihåg att alla skillnader mellan skolorna inte är statistiskt säkerställda enligt konventionella tröskelvärden. Samtidigt är det inte säkert att skillnader måste vara statistiskt signifikanta för att vara av värde i jämförelser. Man behöver inte vara 95 procent säker på att skolor skiljer sig från varandra för att måtten ska vara meningsfulla. Så länge skolor skiljer sig med mer än 50 procents säkerhet är det bättre att använda måttet än att inte använda sig av det. Hur som helst finns det ingenting som hindrar att man kan presentera ovanstående beräknade förädlingsvärden med en specifik felmarginal om man så vill.

7.1.2 Hur bra är Salsa?

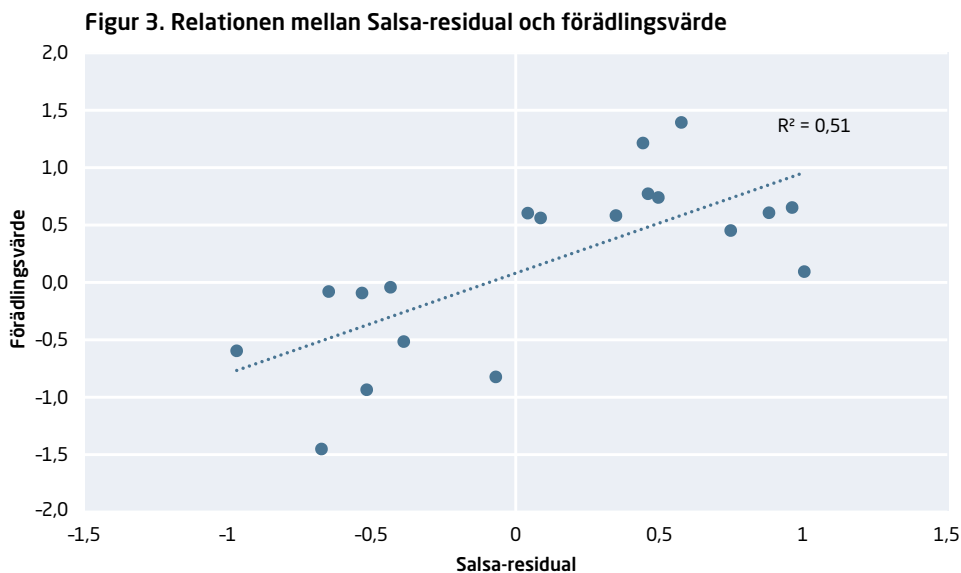
Som beskrevs i avsnitt 3 har forskning visat att förädlingsvärdemått som liknar de som har skapats för kommunens skolor ger en rättvisande bild av skolors och lärares bidrag till inläringen (till exempel Chetty m.fl. 2014a; Deming 2014). En intressant fråga är därför i vilken utsträckning Skolverkets Salsa-modell, som diskuterades i avsnitt 5, överensstämmer med ovanstående resultat. Detta är en viktig fråga eftersom resultatet från denna modell är det enda justerade måttet på skolors kvalitet som finns offentligt tillgängligt. Som beskrevs tidigare i rapporten har föräldrars skolval till årkurs 7 dessutom visat sig påverkas av skolors Salsa-värden när de får denna information hemskickad (Kessel och Olme 2018).

För att jämföra skolornas förädlingsvärde under högstadiet med Skolverkets offentligt tillgängliga Salsa-modell – som beräknas på skolnivå – skapade vi ett Salsa-liknande mått på skolnivå utifrån våra data för kommunens skolor. Skolverkets Salsa-modell använder andelen pojkar, andelen nyinvandrade elever och föräldrarnas högsta genomsnittliga utbildningsnivå (3 kategorier) som förklarande variabler samt skolornas genomsnittliga meritvärde som beroende variabel. Skillnaden mellan skolors faktiska meritvärden och det modellberäknade meritvärdet – variationen i meritvärde som förklaras av bakgrundsvariablerna – ska då motsvara skolors bidrag till inläringen.

Det är viktigt att notera att Salsa-måttet även kan fånga upp skolors effektivitet i årskurser innan årskurs 6, vilket vårt mått filtrerar bort. Men detta visar också på en av Salsa-värdets svagheter: måttet tar inte hänsyn till att olika skolor har olika stadier i grundskolan och att deras elever därför har gått i respektive skola under olika lång tid innan de får sitt slutbetyg i årskurs 9. Vårt mått tar bättre hänsyn till detta fenomen, eftersom vi håller elevernas resultat när de går i årskurs 6 konstant.²³ Eventuella skillnader mellan förädlingsvärdet och Salsa-värdet fångar alltså delvis sannolikt upp sådana skillnader i skolform.

²³ Vårt mått kan inte ta hänsyn till detta helt och hållet, eftersom vissa elever byter skola även under högstadiet. Men detta sker i skolor som erbjuder undervisning i olika årskurser och dessutom i ganska liten utsträckning.

Eftersom analysen över skolors förädlingsvärden baseras på resultat i nationella prov används skolornas genomsnittliga provresultat i årskurs 9 som beroende variabel, i stället för meritvärde, i vårt Salsa-liknande mått.²⁴ Salsa-residualen beräknas här alltså som skillnaden mellan skolors genomsnittliga resultat på de nationella proven i engelska, matematik och svenska i årskurs 9 och de modellberäknade resultaten. Vi redovisar sambandet mellan skolornas förädlingsvärden och residualen från den Salsa-liknande modellen i en figur med inlagd trendlinje och ett R^2 -värde som visar hur mycket av variationen i förädlingsvärdet som förklaras av Salsa-residualen; se figur 3.



Som figur 3 visar förklarar Salsa-residualen 51 procent av variationen i vårt mått som mäter skolornas förädlingsvärde mellan årskurs 6 och 9. Salsa är sannolikt alltså relativt informativt om dessa skolors bidrag till inläringen, men en hög andel av variationen lämnas samtidigt oförklarad. I analyser där vi jämför Salsa-residualen med förädlingsvärdet, som utöver provresultatet i årskurs 6 även justerats för elevernas bakgrundsvariabler, ökar förklaringsvärdet av Salsa-residualen något till 56 procent. Fortfarande är dock nästan hälften av variationen i skolors bidrag till inläringen under högstadiet enligt förädlingsvärdet oförklarad.

Även om vi här använder resultaten på nationella prov som utfallsmått i Salsa-modellen tyder detta på att den är ett ganska informativt mått och att det därför tycks vara rationellt att åtminstone delvis basera sitt val av skola efter den, givet den information som i dag finns tillgänglig. Samtidigt fångar inte Salsa-värdena upp hela bilden vad gäller skolors bidrag till elevers inläring under högstadiet, enligt de mått på skolornas förädlingsvärden som i dag kan skapas med hjälp av nationella prov. Detta påvisar också vikten av att konstruera relevanta mått på förädlingsvärden i Sverige.

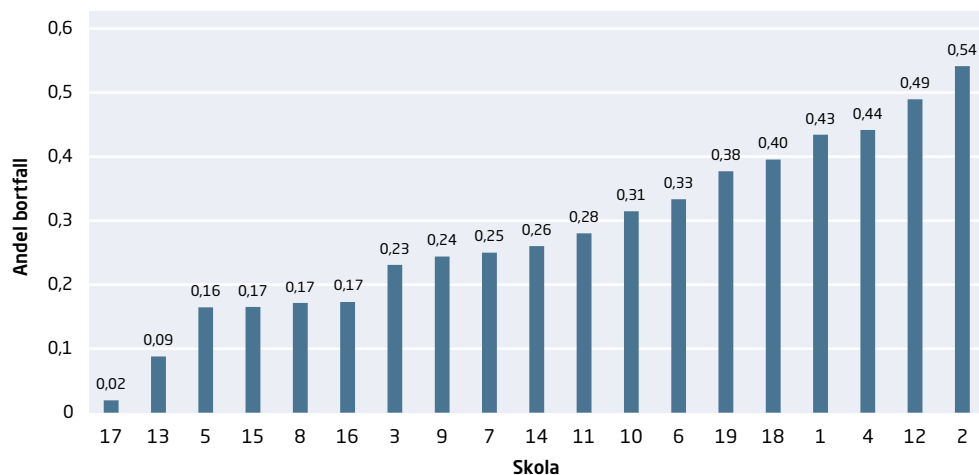
7.1.3 Bortfall är ett problem

Tyvärr visar det sig att vi inte kunde beräkna det genomsnittliga progressionsvärdet för cirka 30 procent av eleverna i kommunen på grund av bortfall på provresultaten i engelska, matematik eller svenska i årskurs 6 eller 9. Som figur 4 visar skiljer sig bortfallet också markant mellan skolorna i kommunen. Skolan med lägst värde har ett bortfall som uppgår till 2 procent, medan skolan med det högsta värdet har ett

²⁴ Genom att använda vårt eget beräknade Salsa-liknande mått kan vi jämföra förädlingsvärdet och ett Salsa-mått som båda bygger på modeller som enbart inkluderar elever i kommunen.

bortfall på 54 procent. Över hälften av skolorna har ett bortfall på över 25 procent, vilket måste anses vara en hög siffra. Med andra ord står det klart att de data vi har att tillgå är långt ifrån perfekta; trots att vi använder data på elevnivå från prov som ska vara obligatoriska är bortfall ett problem för många skolor.

Figur 4. Andel elever för vilka förädlingsvärdet inte kan beräknas på grund av bortfall



Man kan endast spekulera i varför bortfallet är så pass högt som det är. Det kan bero på att ersättningsprov inte rapporteras in i statistiken. Den mest problematiska förklaringen är dock kanske att elever inte dyker upp när de ska skriva provet – eller att skolor med dåliga resultat väljer att inte rapportera in korrekta data. Möjligheterna att göra detta är naturligtvis ganska stora då lärarna själva rättar proven och det inte finns någon egentlig kontrollmekanism för att säkerställa att alla prov som eleven har skrivit faktiskt lämnas in. Enligt lag ska nationella prov vara obligatoriska – men enligt dessa resultat verkar de faktiskt inte vara det.

Det är värt att notera att det inte finns någon relation mellan skolornas förädlingsvärde och deras bortfall – det verkar alltså inte som att skolor uppnår högre förädlingsvärden för att de har högre bortfall. Men det är helt klart anmärkningsvärt att den officiella statistiken visar att över hälften av skolorna i den svenska kommunen som här studeras har ett bortfall på 25 procent eller mer.

7.1.4 Slutsatser från analysen

Sammantaget exemplifierar analysen hur man skulle kunna gå tillväga för att skapa mått på skolors förädlingsvärden vad gäller resultaten i nationella prov i högstadiet. Analysen visar också att Salsa-modellen tycks fånga upp högstadieskolors kvalitet i ganska hög utsträckning, men att den sannolikt också missar viktiga skillnader som existerar mellan skolor. Även om vi har använt nationella prov som utfallsmått skulle man lika gärna kunna använda metoden för att beräkna skolors förädlingsvärden baserade på de nationella proven i matematik och svenska i årskurs 3 och 6 eller betygen i årskurs 6 och 9. Likaså skulle man kunna mäta förädlingsvärdet under gymnasieåren, då man kanske också bör inkludera kontroller för vilken grundskola eleverna gick på samt vilket program de läser på. Som noteras i avsnitt 8.8 är det viktigt att man noga jämför olika specifikationer och metoder innan man skapar nationella mått över skolors förädlingsvärden, för att maximera möjligheten att man lyckas skapa så rigorösa mått som möjligt.

Ett oroväckande resultat som framkom i analysen var att man inte kunde beräkna ett förädlingsvärde för en förvånansvärt hög andel elever, på grund av bortfall i provstatistiken, vilket indikerar att de nationella proven faktiskt inte är obligatoriska i praktiken. Detta leder oss in på frågan om vad som krävs för att skapa ett rättssäkert system för att utvärdera skolors och lärares effektivitet, vilket diskuteras i nästa avsnitt.

8. Förslag på hur man skapar mer rigorösa mått på förädlingsvärden

Man kan alltså utforma mått på skolors förädlingsvärden inom ramen för dagens system. Samtidigt står det också klart att det finns stora brister som gör att de data som måtten baseras på långt ifrån är fullgoda. Nedan diskuteras därför vad som bör göras för att kunna skapa så rigorösa mått på förädlingsvärden i skolsystemet som möjligt. Förslagen är fristående från varandra och behöver inte nödvändigtvis kombineras för att vara av värde. Vissa är specifika för grundskolan, medan andra även gäller gymnasieskolan.

8.1 Rätta de nationella proven externt och minska möjligheterna till fusk

Ett av de viktigaste stegen för att kunna förbättra informationen om skolors kunskapsmässiga kvalitet vore att införa extern och anonym rättning på de nationella proven. Utan denna grundbult blir det svårt att skapa ett rättssäkert system för att kunna mäta skolors förädlingsvärden i denna bemärkelse. Detta eftersom betygssättningen av nationella prov i dag inte är likvärdig, vilket tydligt påvisas i Skolinspektionens (2018) studier. Extern rättning kan säkerställas genom att alla prov anonymiseras och skickas till andra skolor eller huvudmän. Eftersom lärarna i dag rättar sina egna elevers prov skulle arbetsbördan vara densamma och systemet kommer också underlättas av den kommande digitaliseringen av de nationella proven. Alla lärare och huvudmän vet att de rättar sina konkurrenters resultat och alla har därför starkare incitament att hålla nere betygen än att inflatera dem, vilket inte är fallet i dag. Alternativt kan man göra som i England, där de olika examinationsorganisationerna helt enkelt betalar lärare för att rätta proven.

För att minimera sannolikheten att fusk och läckor förekommer bör proven hanteras digitalt av Skolverket och genomföras samtidigt över hela landet. Likaså bör man utreda huruvida själva provskrivningen bäst övervakas av oberoende skrivvakter, ungefär som i dag sker vid högskoleprovet, i dag agerar lärare och andra personer skrivvakter på sina egna skolor; med det föreslagna systemet skulle man i stället övervaka grannskolans skrivningar. På så sätt skulle arbetsbördan vara densamma. Skillnaden är i stället att man maximerar möjligheterna att reglerna faktiskt efterlevs.

8.2 Säkerställ att de nationella proven skrivs och att resultaten rapporteras in

Likaså måste det klargöras att de nationella proven är obligatoriska. Även om det endast går att kommentera de data som finns från kommunen som studeras i avsnitt 7 är bortfallet för flera av denna kommuns skolor oroväckande högt. Det är därför värt att i varje fall utreda huruvida resultaten bland elever som helt enkelt struntar i att skriva proven över huvud taget – utan godkända anledningar – bör rapporteras in som att de får betyget F i stället för att de är frånvarande. Myndigheter måste också följa upp så att alla elevers resultat faktiskt rapporteras in korrekt. Detta förfarande bör underlättas av att hanteringen av proven sker digitalt.

8.3 Säkerställ att man kan beräkna förädlingsvärden redan från första klass

Det krävs flera andra förändringar för att mäta förädlingsvärden mer generellt i grundskolan. Vi är exempelvis naturligtvis intresserade av att kunna mäta skolors effektivitet även i lågstadiet. Detta kräver i sin tur att elever skriver ett diagnostiskt prov precis innan de börjar första klass – eller precis efter att de har börjat – och att resultaten från detta prov samlas in. Resultaten från dessa prov kan sedan kombineras med resultaten från de nationella proven i årskurs 3 samt de nationella proven i årskurs 6. Även de diagnostiska proven bör i första klass rättas externt för att säkerställa att betygssättningen är likvärdig.

8.4 Skapa mått som baseras på årliga diagnostiska prov med relativ rättning

Att skapa förädlingsvärden efter resultat på nationella prov innebär att man använder kunskapskontroller som elever och lärare har starka incitament att prestera så bra som möjligt på. Samtidigt är det inte nödvändigt att detta är en fördel när det gäller utformningen av förädlingsvärden. 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 över huvud taget. Eftersom de kunskapskontroller som används generellt måste utformas på ett likartat sätt över tid, finns det alltid möjligheter att träna elever för att klara av dem i stället för att lära ut bredare kunskaper. Om man knyter starka incitament till dessa kunskapskontroller finns det alltså risk att deras prediktiva värde för framtida utfall minskar.

Ett alternativ är därför att i stället skapa förädlingsvärdesmått baserade på prov som inte spelar någon roll för elevernas faktiska resultat, utan endast används för att följa kunskapsutvecklingen (se Heller-Sahlgren och Jordahl 2016). Ett sätt är att knyta incitament bland lärare och skolor till resultat som baseras på relativa prestationer, en idé som presenteras av Neal (2010) och Barlevy och Neal (2012). I denna modell används diagnostiska prov med relativ rättning för att jämföra prestationer bland elever med liknande bakgrund och initiala resultat. Eftersom fokus endast ligger på relativa prestationer kan formatet på proven, och till viss del innehållet, variera mellan olika år. Det gör det svårt för skolorna och lärarna att helt enkelt coacha eleverna för specifika prov. Istället skulle ett sådant system tvinga fram fokus på det som examineras som helhet snarare än endast det som mäts med ett specifikt provformat. Detta eftersom skolorna och lärarna inte vet vad för slags prov som eleverna måste skriva; därmed måste de lära ut sådant som krävs för att elever ska göra bättre ifrån sig på alla slags prov.

Relativt rättade prov har alltså fördelen att deras utformning kan ändras över tid. Ibland kallas dessa prov normrelaterade. Om förädlingsvärdena baseras på sådana prov bör de vara helt frikopplade från betygssystemet. Det minskar möjligheter och incitament bland skolor att manipulera resultaten. Eftersom dessa diagnostiska prov frikopplas från betygen har de i sin tur ingen betydelse för elevernas framtid annat än för att ge återkoppling till lärare om hur elevers relativa resultat förändras över tid samt för att hålla skolor ansvariga gentemot både föräldrar och myndigheter (och lärare ansvariga gentemot sina rektorer).

På så sätt kan man alltså tänka sig att elever skriver årliga diagnostiska prov – vars format förändras över tid – i åtminstone ett par ämnen. Med hjälp av årliga relativa prov förbättras möjligheterna till uppföljning och insatser markant, liksom förädlingsvärdenas stabilitet, samtidigt som det föränderliga provformatet minskar möjligheterna att manipulera resultaten.

8.5 Mät förädlingsvärden som baseras på andra utfall än endast prov

Som beskrevs i avsnitt 3 finns det andra färdigheter än de som fångas upp i prov som har stor betydelse för elevers framtid. Det går i dag inte att skapa mått på skolors eller lärares förädlingsvärden när det gäller icke-kognitiva utfall i sig. Detta eftersom data vad gäller mått som ämnar fånga upp elevers icke-kognitiva utfall – såsom frånvaro – inte rapporteras in till myndigheterna.

För att kunna skapa mått på skolors effektivitet vad gäller renodlade icke-kognitiva färdigheter skulle det alltså krävas att data för relevanta utfall börjar rapporteras in till myndigheterna. Det kan dock finnas problem med att använda sådana mått, eftersom de i hög grad bygger på att skolor och lärare själva rapporterar in data över frånvaro. Å andra sidan är det precis så systemet för rapporteringen för betyg och resultat på nationella prov fungerar i dag, vilket inte hindrar att dessa data offentliggörs och diskuteras. Dessutom behöver man inte nödvändigtvis offentliggöra förädlingsvärden vad gäller icke-kognitiva utfall för att de ska vara användbara för myndigheter, rektorer och lärare.

Om man inte vill gå så långt som att samla in nya data vad gäller elevers icke-kognitiva utfall är ett alternativ att i stället skapa förädlingsvärden som enbart baseras på elevers genomsnittliga betyg, vilket forskningen som diskuterades i avsnitt 3 indikerar fångar upp den totala effekten av lärares icke-kognitiva förädlingsvärden relativt väl. Eftersom även betygen i dag bestäms lokalt bör dessa dock också i sådana fall göras rättssäkra, exempelvis genom att man förankrar dem i de nationella proven som rättas externt så att skolornas medelbetyg inte får överstiga deras snittresultat på dessa prov. På så sätt skulle man även kunna tygla eventuell betygsinflation. Genom att kombinera förädlingsvärdena som baseras på prov med de som baseras på betyg skulle det prediktiva värdet sannolikt öka markant, i enlighet med den forskning som diskuterades i avsnitt 3.

8.6 Samla in alla tillgängliga betyg i slutet av varje läsår

I dag samlas inte betygen i årskurs 7 och 8 in över huvud taget, medan kursbetygen i årskurs 1 och 2 på gymnasiet endast samlas in i slutet av gymnasieutbildningen. Som diskuterades i avsnitt 6.1 mäter den norska motsvarigheten till Skolverket numera gymnasieskolors förädlingsvärden i varje årskurs för olika program. För att kunna beräkna liknande värden i Sverige och uppdatera resultaten så att de finns tillgängliga så snart som möjligt krävs att man börjar samla in dessa data årligen. Om man även börjar samla in betygen i årskurs 7 och 8 i grundskolan med samma tidsintervall skulle man också kunna mäta förädlingsvärden vad gäller betyg årligen på högstadiet.

Detta förslag har också bäring på måttens kvalitet, eftersom man kan reducera mätfel och öka deras prediktiva värde om man har tillgång till data från fler tillfällen. För att kunna skapa så relevanta förädlingsvärden som möjligt bör därför alla tillgängliga betyg samlas in vid läsårets slut.

8.7 Säkerställ att man kan koppla lärare till elever i nationella databaser

Som har diskuterats i tidigare avsnitt har amerikanska skolmyndigheter valt att beräkna och publicera enskilda lärares förädlingsvärden i stället för att endast fokusera på skolors effektivitet. Detta går inte att göra i Sverige eftersom det i de data som samlas in av statliga myndigheter inte går att avgöra vilka lärare som undervisar vilka elever.²⁵ Man kan helt enkelt inte mäta lärares effektivitet på nationell nivå, eftersom det inte går att koppla elever till de lärare som har undervisat dem.

²⁵ Vissa kommuner samlar in data som gör att man kan matcha elever till lärare (se Grönqvist och Vlachos 2016), men man kan för tillfället inte beräkna förädlingsvärden för landets alla lärare.

För att man ska kunna beräkna enskilda lärares förädlingsvärden krävs alltså att man börjar samla in denna information systematiskt. På så sätt skulle man också kunna ta fram mått som kan användas vid anställningsförfaranden och för skolutvecklingssyften. När man sedan har tillgång till dessa data kan man utforma måtten på förädlingsvärden bland lärare på samma sätt som man utformar dem bland skolor. Huruvida man bör offentliggöra enskilda lärares förädlingsvärden är dock en fråga som kräver mer utredning.

8.8 Tillsätt en expertgrupp för att utreda exakt hur måtten bör utformas

Som beskrevs i avsnitt 6.1 föregicks den norska publiceringen av gymnasieskolors förädlingsvärden av forskning som analyserade hur måtten kunde och borde utformas. Det norska skolverket baserade sedan utformningen av förädlingsvärdena i mångt och mycket på denna forskning. Bidraget som redovisas i avsnitt 7 är ett försök att visa hur man kan beräkna specifika förädlingsvärden i den svenska kontexten, men innan man bestämmer sig för exakt hur måtten bör utformas och spridas nationellt rekommenderar jag att regeringen tillsätter en expertgrupp av forskare med uppdrag att utforma och analysera olika slags mått på skolors och lärares förädlingsvärden, både vad gäller provresultat och betyg samt eventuellt också icke-kognitiva utfall i den svenska kontexten.

Man bör exempelvis utreda huruvida svenska lärare och skolor är olika effektiva för olika slags elever. I England har man funnit att skolor är olika effektiva för elever med olika initiala resultat (Dearden m.fl. 2011). Man har därför publicerat måtten uppdelade efter elevers initiala resultat för att ta hänsyn till dessa skillnader.²⁶ Det vore därför önskvärt att man åtminstone utreder huruvida svenska lärare och skolor är olika effektiva för elever med olika bakgrund. Om så är fallet bör måtten som tas fram också reflektera detta.

Genom att jämföra i vilken utsträckning olika mått fångar upp lärares och skolors effektivitet, deras prediktiva värde för långsiktiga utfall samt vilka möjligheter till manipulation de skapar skulle man maximera möjligheterna att skapa så rigorösa mått som möjligt.

²⁶ I Sverige har forskning som använder data från ett par svenska kommuner visat att lärare med olika kognitiva och icke-kognitiva förmågor är olika effektiva för olika slags elever (Grönqvist och Vlachos 2016).

9. Slutsatser

Den här rapporten har diskuterat vikten av ett välfungerande system för kvalitetsinformation som baseras på förädlingsvärden – skolors och lärares faktiska bidrag till elevernas inläring och andra relevanta utfall – samt hur förädlingsvärden i den svenska skolan skulle kunna utformas. Den har dragit lärdomar från forskningen kring effekterna av sådana kvalitetsmått och även diskuterat hur andra länder har utformat relevanta system.

Rapporten har visat att lärares och skolors förädlingsvärden kan utformas så att de fångar upp deras bidrag till elevers inläring och framtida utfall. System som mäter och offentliggör skolors och lärares förädlingsvärden har dessutom ofta haft positiva effekter på elevers inläring. Samtidigt står det klart att lärare påverkar viktiga utfall bland eleverna som inte nödvändigtvis fångas upp i deras förädlingsvärden enligt provresultat. Detta indikerar att det är viktigt att mäta lärares och skolors effektivitet både på prov och enligt andra utfall.

Tyvärr står det klart att det svenska utbildningssystemet länge saknat ett ordentligt system för information om skolors kvalitet – och de kvalitetsmått som existerar är inte tillräckligt robusta. Situationen i Sverige skiljer sig också från länder som Norge, England och USA, där man i dag systematiskt mäter (och ofta offentliggör) förädlingsvärden i skolsystemet. I USA mäter flera delstater och skoldistrikt även enskilda lärares förädlingsvärden.

Inom ramen för dagens svenska system kan man skapa förädlingsvärden på skolnivå från och med årskurs 3, då elever för första gången skriver nationella prov i matematik och svenska. I årskurs 6 skriver elever sedan för första gången nationella prov i alla tre kärnämnen och får även betyg i en rad ämnen, som dock varierar mellan olika skolor. Rapporten visade också hur man kan gå tillväga för att skapa mått på skolors förädlingsvärden vad gäller resultaten i nationella prov i högstadiet, med data på elevnivå från en svensk kommun. Analysen visade att modellen för det enda justerade kvalitetsmått som existerar nationellt – Salsa – tycks fånga upp högstadieskolors kvalitet i ganska hög utsträckning, men att den också missar viktiga skillnader. Ett oroväckande resultat är samtidigt att man inte kan beräkna ett förädlingsvärde för en förvånansvärt hög andel elever, på grund av bortfall i provstatistiken.

För att kunna skapa bättre kvalitets- och informationsmått i Sverige krävs ett antal åtgärder. Rapporten gav därför förslag på sådana åtgärder:

1. Rätta de nationella proven externt och minska möjligheterna till fusk

Utan extern och anonym rättning blir det svårt att skapa ett rättssäkert system för att kunna mäta skolors förädlingsvärden på nationella prov. För att minimera sannolikheten att fusk förekommer bör proven hanteras digitalt av Skolverket, medan provskrivningen bör övervakas av oberoende skrivvakter. Alla prov bör också skrivas samtidigt över hela landet.

2. Säkerställ att de nationella proven skrivs och att resultaten rapporteras in

Man måste säkerställa att alla elever faktiskt skriver de nationella proven. Myndigheter måste också följa upp så att alla elevers resultat rapporteras in korrekt. Detta förfarande bör underlättas av att hanteringen av proven sker digitalt.

3. **Säkerställ att man kan beräkna förädlingsvärden redan från första klass**
För att kunna mäta förädlingsvärden från första klass bör elever också skriva ett diagnostiskt prov precis innan de börjar första klass – eller precis efter att de har börjat – och resultaten från detta prov bör samlas in.
4. **Skapa mått som baseras på årliga diagnostiska prov med relativ rättning**
Med hjälp av årliga normrelaterade prov förbättras möjligheterna till uppföljning och insatser markant, liksom förädlingsvärdenas stabilitet, samtidigt som man kan förändra provformatet för att på så sätt minska möjligheterna att manipulera resultaten.
5. **Mät förädlingsvärden som baseras på andra utfall än endast prov**
För att kunna skapa mått på förädlingsvärden vad gäller bredare utfall krävs att data för sådana utfall – såsom frånvaro – börjar rapporteras in till myndigheterna. Ett enklare alternativ vore dock att skapa förädlingsvärden som baseras på elevers genomsnittliga betyg, vilket forskningen indikerar fångar upp den totala effekten av lärares icke-kognitiva förädlingsvärde relativt väl. Dessa betyg bör samtidigt förankras i de nationella proven som rättas externt, så att skolornas medelbetyg inte får överstiga deras snittresultat på dessa prov.
6. **Samla in alla tillgängliga betyg i slutet av varje läsår**
Den norska motsvarigheten till Skolverket beräknar gymnasieskolors förädlingsvärden i varje årskurs för olika program. För att kunna beräkna liknande värden i Sverige inom rimlig tid krävs att man börjar samla in dessa data årligen. Om man även börjar samla in betygen i årskurs 7 och 8 i grundskolan med samma tidsintervall skulle man också kunna mäta förädlingsvärden vad gäller betyg årligen på högstadiet.
7. **Säkerställ att man kan koppla lärare till elever i nationella databaser**
Man kan i dag inte mäta lärares effektivitet på nationell nivå, eftersom det inte går att koppla elever till de lärare som har undervisat dem. För att man ska kunna beräkna enskilda lärares förädlingsvärden krävs att man börjar samla in denna information systematiskt.
8. **Tillsätt en expertgrupp av forskare för att utreda exakt hur måtten bör utformas**
Genom att jämföra i vilken utsträckning olika mått fångar upp lärares och skolors effektivitet, deras prediktiva värde för långsiktiga utfall samt vilka möjligheter till manipulation de skapar skulle man maximera möjligheterna att skapa så rigorösa mått som möjligt. Man bör exempelvis utreda huruvida svenska lärare och skolor är olika effektiva för olika slags elever. Om så är fallet bör måtten som tas fram också reflektera detta.

Ovanstående förslag skulle avsevärt underlätta möjligheterna att skapa rigorösa mått på skolors och lärares förädlingsvärden i Sverige. Sådana mått skulle i sin tur ge starkare incitament till alla aktörer i skolsystemet att dra åt samma håll i jakten på ökade kunskaper och färdigheter – samt ge dem bättre verktyg för att de ska kunna bli framgångsrika i denna bemärkelse.

Referenser

- Abdulkadiroglu, Atila, Parag A. Pathak, Jonathan Schellenberg och Christopher R. Walters. 2017. “Do Parents Value School Effectiveness?” NBER Working Paper No. 23912, National Bureau of Economic Research, Cambridge, MA.
- Adnot, Melinda, Thomas Dee, Veronica Katz och James Wyckoff. 2017. “Teacher Turnover, Teacher Quality och Student Achievement in DCPS.” *Educational Policy and Policy Analysis* 39(1):54–76.
- Ajayi, Kehinde F. 2013. “School Choice and Educational Mobility: Lessons from Secondary School Applications in Ghana.” Manuskript, Boston University, Boston.
- Ajayi, Kehinde F., Willa H. Friedman och Adrienne M. Lucas. 2017. “The Importance of Information Targeting for School Choice.” *American Economic Review: Papers & Proceedings* 107(5):638–643.
- Andrabi, Tahir, Jishnu Das och Asim I. Khwaja. 2017. “Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets.” *American Economic Review* 107(6):1535–1563.
- Angrist, Joshua D., Peter Hull, Parag Pathak och Christopher Walters. 2016. “Interpreting Tests of School VAM Validity.” *American Economic Review: Papers & Proceedings* 106(5):388–392.
- Angrist, Joshua D., Peter D. Hull, Parag A. Pathak och Christopher R. Walters. 2017. “Leveraging Lotteries for School Value-added Testing and Estimation.” *Quarterly Journal of Economics* 132(2):871–919.
- Arevik, Niklas. 2014. “‘Förädlingsvärde’ ska underlätta skolvalet.” *Lärarnas tidning*, januari 10:e, pp. <https://lararnastidning.se/foradlingsvarde-ska-underlatta-skolvalet/>.
- Autor, David, Marianne Bertrand, Esther Duflo, Benjamin Feigenberg och Francisco Gallego. 2016. “Estimating supply and demand responses to an information intervention in the Chilean education sector.” 3ie Grantee Final Report, International Initiative for Impact Evaluation (3ie), New Delhi.
- Bacher-Hicks, Andrew, Mark J. Chin, Thomas J. Kane och Douglas O. Staiger. 2017. “An Evaluation of Bias in Three Measures of Teacher Quality: Value-Added, Classroom Observations, and Student Surveys.” NBER Working Paper No. 23478, National Bureau of Economic Research, Cambridge, MA.
- Bacher-Hicks, Andrew, Thomas J. Kane och Douglas O. Staiger. 2014. “Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles.” NBER Working Paper No. 20657, National Bureau of Economic Research, Cambridge, MA.
- Ballou, Dale och Matthew G. Springer. 2017. “Has NCLB Encouraged Educational Triage? Accountability and the Distribution of Achievement Gains.” *Education Finance and Policy* 12(1):77–106.
- Barlevy, Gadi och Derek Neal. 2012. “Pay for Percentile.” *American Economic Review* 102(5):1805–1831.

- Bergman, Peter och Matthew J. Hill. 2018. "The effects of making performance information public: Regression discontinuity evidence from Los Angeles teachers." *Economics of Education Review* 66:104–113.
- Bifulco, Robert. 2012. "Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison." *Journal of Policy Analysis and Management* 31(3):729–751.
- Blazar, David och Matthew A. Kraft. 2017. "Teacher and Teaching Effects on Students' Attitudes and Behaviors." *Educational Evaluation and Policy Analysis* 39(1):146–170.
- Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman och John E. Humphries. 2016. "What grades and achievement tests measure." *PNAS* 113(47):13354–13359.
- Brasington, David M. och Donald R. Haurin. 2006. "Educational Outcomes and House Values: A Test of the Value-Added Approach." *Journal of Regional Science* 46(2):245–268.
- Buddin, Richard. 2010. "How effective are Los Angeles elementary teachers and schools?" MPRA Paper No. 27366, Munich Personal RePEc Archive.
- 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.
- Camargo, Braz, Rafael Camelo, Sergio Firpo och Vladimir Ponczek. 2018. "Information, Market Incentives, and Student Performance: Evidence from a Regression Discontinuity Design in Brazil." *Journal of Human Resources* 53(2):414–444.
- Carlson, Deven E., Joshua M. Cowen och David J. Fleming. 2013. "Third-Party Governance and Performance Measurement: A Case Study of Publicly Funded Private School Vouchers." *Journal of Public Administration Research and Theory* doi:10.1093/jopart/mut017.
- Carlson, Deven och Stéphane Lavertu. 2016. "Charter school closure and student achievement: Evidence from Ohio." *Journal of Urban Economics* 95:31–48.
- Cedefop. 2015. "Skill Supply and Demand Up to 2025." Rapport, Thessaloniki. <http://www.cedefop.europa.eu/printpdf/publications-and-resources/country-reports/sweden-skills-forecasts-2025>.
- Chakrabarti, Rajashri. 2014. "Incentives and Responses under No Child Left Behind: Credible Threats and the Role of Competition." *Journal of Public Economics* 110:124–146.
- Chetty, Raj, John F. Friedman och Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers 1: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104(9):2593–2632.
- Chetty, Raj, John N. Friedman och Jonah E. Rockoff. 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104(9):2633–2679.
- Chetty, Raj, John N. Friedman och Jonah E. Rockoff. 2016. "Using Lagged Outcomes to Evaluate Bias in Value-Added Models." *American Economic Review Papers & Proceedings* 106(5):393–399.
- Chetty, Raj, John N. Friedman och Jonah E. Rockoff. 2017. "Measuring the Impacts of Teachers: Reply." *American Economic Review* 107(6):1685–1717.

- Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics* 93:1045–1057.
- Coe, Rob. 2014. "Classroom observation: it's harder than you think." Blogg, Centre for Evaluation & Monitoring, Durham University. <http://www.cem.org/blog/414/>.
- Cullen, Julie B., Brian A. Jacob och Steven A. Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica* 74(5):1191–1230.
- Cullen, Julie B., Corry Koedel och Eric Parsons. 2017. "The Compositional Effect of Rigorous Teacher Evaluation on Workforce Quality." CALDER Working Paper No. 168, National Center for Analysis of Longitudinal Data in Education Research, Washington, DC.
- DCSF. 2003a. "Technical annex." Teknisk rapport om skolors förädlingsvärden mellan årskurs 6 och 11, mellan årskurs 6 och 9 samt mellan årskurs 9 och 11, Department for Children, Schools and Families, London. https://webarchive.nationalarchives.gov.uk/20150812154312/http://www.education.gov.uk/schools/performance/archive/va1_03/docD.shtml.
- DCSF. 2003b. "Value added technical information." Teknisk rapport om skolors förädlingsvärden mellan årskurs 2 och 6, Department for Children, Schools and Families, London. https://webarchive.nationalarchives.gov.uk/20150812154104/http://www.education.gov.uk/schools/performance/archive/primary_03/p5.shtml.
- DCSF. 2007. "A technical guide to contextual value added: 2007 model." Department for Children, Schools and Families, London. https://webarchive.nationalarchives.gov.uk/20150812152000/http://www.education.gov.uk/schools/performance/archive/schools_07/s3.shtml.
- DCSF. 2009. "Publication of 2009 test and examination results in the school and college achievement and attainment tables. Statement of intent – 2009." Department for Children, Schools and Families, London. <https://webarchive.nationalarchives.gov.uk/20150812113237/http://www.education.gov.uk/schools/performance/archive/2009-statement-of-intent-v1-0.pdf>.
- Dearden, Lorraine, John Micklewright och Anna Vignoles. 2011. "The Effectiveness of English Secondary Schools for Pupils of Different Ability Levels." *Fiscal Studies* 32(2):225–244.
- Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review Papers & Proceedings* 104(5):406–411.
- Deming, David J., Sarah Cohodes, Jennifer Jennings och Christopher Jencks. 2016. "School Accountability, Postsecondary Attainment, and Earnings." *Review of Economics and Statistics* 98(5):848–862.
- Deming, David J., Justine S. Hastings, Thomas J. Kane och Douglas O. Staiger. 2014. "School Choice, School Quality, and Postsecondary Attainment." *American Economic Review* 104(3):991–1013.
- DfE. 2010. "The importance of teaching: The Schools White Paper 2010." Rapport, London.
- DfE. 2018a. "Primary school accountability in 2018: A technical guide for primary maintained schools, academies and free schools." Department for Education, London.

- DfE. 2018b. "Secondary accountability measures Guide for maintained secondary schools, academies and free schools." Department for Education, London.
- DfE. 2018c. "16 to 18 accountability measures: technical guide." Department for Education, London.
- DfES. 2004. "Autumn Package 2004." Department for Education and Skills, London. <https://webarchive.nationalarchives.gov.uk/20050301213057/http://www.standards.dfes.gov.uk/performance/ap/>.
- Downes, Thomas A. och Jeffrey E. Zabel. 2002. "The Impact of School Characteristics on House Prices: Chicago 1987–1991." *Journal of Urban Economics* 52(1):1–25.
- Edin, Per-Anders, Peter Fredriksson, Martin Nybom och Björn Öckert. 2017. "The Rising Return to Non-Cognitive Skill." IZA Discussion Paper No. 10914, IZA Institute of Labor Economics, Bonn.
- Falch, Torberg, Simon Bensnes och Bjarne Strøm. 2016. "Skolekvalitet i videregående opplæring – Utarbeidelse av skolebidragsindikatorer og mål på skolekvalitet." SØF-rapport nr. 01/16, Senter for økonomisk forskning, Trondheim.
- Ferrão, Maria E. 2012. "On the Stability of Value Added Indicators." *Quality & Quantity* 46:627–637.
- Figlio, David och Susanna Loeb 2011. "School Accountability." Pp. 383–421 in *Handbook of the Economics of Education*, redigerad av Eric A Hanushek, Stephen Machin och Ludger Wößmann. Amsterdam: North-Holland.
- Fortson, Kenneth, Natalya Verbitsky-Savitz, Emma Kopa och Philip Gleason. 2012. "Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods Can Replicate Experimental Impact Estimates." Rapport, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences (IES), Washington, DC.
- Fuller, Sarah C. och Helen F. Ladd. 2012. "School Based Accountability and the Distribution of Teacher Quality Among Grades in Elementary Schools." Paper prepared for Association for Education Finance and Policy, Boston, MA.
- Gallego, Francisco A., Francisco Lagos och Yael Stekel. 2012. "Types of Information and School Choice: An Experimental Study in the Chilean Voucher System." Manuskript, Pontificia Universidad Católica de Chile, Santiago.
- Gershenson, Seth. 2016a. "Performance Standards and Employee Effort: Evidence From Teacher Absences." *Journal of Policy Analysis and Management* 35(3):615–638.
- Gershenson, Seth. 2016b. "Linking Teacher Quality, Student Attendance, and Student Achievement." *Education Finance and Policy* 11(2):125–149.
- Gibbons, Stephen, Stephen Machin och Olmo Silva. 2013. "Valuing School Quality Using Boundary Discontinuities." *Journal of Urban Economics* 75:15–28.
- Gibbons, Stephen och Olmo Silva. 2011. "School Quality, Child Wellbeing, and Parents' Satisfaction." *Economics of Education Review* 30:312–331.
- Glazerman, Steven och Ali Protik. 2015. "Validating Value-added Measures of Teacher Performance." Konferensrapport, Annual meeting of the American Economic Association den 5:e januari 2015.

- Grönqvist, Erik och Jonas Vlachos. 2016. “One size fits all? The effects of teachers’ cognitive and social abilities on student achievement.” *Labour Economics* 42:138–150.
- Hagström, Per. 2015. “Farligt mäta skolors förädlingsvärde.” *Lärarnas tidning*, augusti 26:e, pp. <https://lararnastidning.se/farligt-mata-skolors-foradlingsvarde/>.
- Hastings, Justine S. och Jeffrey M. Weinstein. 2008. “Information, School Choice, and Academic Achievement: Evidence from Two Experiments.” *Quarterly Journal of Economics* 123(4):1373–1414.
- Haunshek, Eric A. och Margaret E. Raymond. 2005. “Does School Accountability Lead to Improved Student Performance?” *Journal of Policy Analysis and Management* 24(2):297–327.
- Hayek, Friedrich A. 1945. “The Use of Knowledge in Society.” *American Economic Review* 35(4):519–30.
- Heller-Sahlgren, Gabriel. 2017. “Skolpolitiskt relevanta faktorer bakom Sveriges resultat i Pisa 2015.” Rapport, Svenskt Näringsliv, Stockholm.
- Heller-Sahlgren, Gabriel. 2019. “Education Policy for Health Equality: Lessons for the Nordic Region.” Rapport, Nordic Welfare Centre, Stockholm.
- Heller-Sahlgren, Gabriel och Henrik Jordahl. 2016. *Information – ett verktyg för bättre skolsystem*. Stockholm: SNS Förlag.
- Hubbard, Daniel. 2018. “More Gains Than Score Gains? High School Quality and College Success.” Job Market Paper, Department of Economics, University of Michigan, Ann Arbor, MI.
- Imberman, Scott A. och Michael F. Lovenheim. 2016. “Does the Market Value Value-Added? Evidence from Housing Prices After a Public Release of School and Teacher Value-Added.” *Journal of Urban Economics* 91:104–121.
- Jackson, C. K. 2018. “What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes.” *Journal of Political Economy* 2072–2107.
- Jackson, C. K., Jonah E. Rockoff och Douglas O. Staiger. 2014. “Teacher Effects and Teacher-Related Policies.” *Annual Review of Economics* 6:801–825.
- Jacob, Brian A. och Lars Lefgren. 2008. “Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education.” *Journal of Labor Economics* 26(1):101–136.
- Kane, Thomas J., Daniel F. McCaffrey, Trey Miller och Douglas O. Staiger. 2013. “Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment.” Rapport, The Measures of Effective Teaching (MET) project.
- Kane, Thomas J. och Douglas O. Staiger. 2008. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” NBER Working Paper No. 14607, National Bureau of Economic Research, Cambridge, MA.
- Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel och Lex Borghans. 2017. “Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success.” NBER Working Paper No. 20749, National Bureau of Economic Research, Cambridge, MA.

- Kessel, Dany och Elisabeth Olme. 2018. "Are Parents Uninformed? The Impact of School Performance Information on School Choice Behaviour and Student Assignment." Avsnitt i Dany Kessels doktorsavhandling "School Choice, School Performance and School Segregation: Institutions and Design", Stockholms universitet, Stockholm.
- Koedel, Cory, Kata Mihaly och Jonah E. Rockoff. 2015. "Value-added modeling: A review." *Economics of Education Review* 47:180–195.
- Koning, Pierre och Karen van der Wiel. 2012. "School Responsiveness to Quality Rankings: An Empirical Analysis of Secondary Education in the Netherlands." *De Economist* 160(4):339–355.
- 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.
- Lai, Fang, Elisabeth Sadoulet och Alain de Janvry. 2009. "The adverse effects of parents' school selection errors on academic achievement: Evidence from the Beijing open enrollment program." *Economics of Education Review* 28(4):485–496.
- Lauen, Douglas L. 2013. "Jumping at the Chance: The Effects of Accountability Incentives on Student Achievement." *Journal of Research on Educational Effectiveness* 6(2):93–113.
- Lauen, Douglas L. och Michael S. Gaddis. 2012. "Shining a Light or Fumbling in the Dark? The Effects of NCLB's Subgroup-Specific Accountability on Student Achievement." *Educational Evaluation and Policy Analysis* DOI: 10.3102/0162373711429989.
- 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.
- Leckie, George och Harvey Goldstein. 2009. "The Limitations of Using School League Tables to Inform School Choice." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 172(4):835–851.
- Leckie, George och Harvey Goldstein. 2017. "The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'." *British Educational Research Journal* 43(2):193–212.
- Lefgren, Brian A. och Lars Lefgren. 2007. "What Do Parents Value in Education? An Empirical Investigation of Parents' Revealed Preferences for Teachers." *Quarterly Journal of Economics* 122(4):1603–1637.
- Lucas, Adrienne M. och Isaac M. Mbiti. 2012. "The Determinants and Consequences of School Choice Errors in Kenya." *American Economic Review* 102(3):283–288.
- Macartney, Hugh. 2016. "The Dynamic Effects of Educational Accountability." *Journal of Labor Economics* 34(1):1–28.
- Macartney, Hugh, Robert McMillan och Uros Petronijevic. 2015. "Incentive Design in Education: An Empirical Analysis." NBER Working Paper No. 21835, National Bureau of Economic Research, Cambridge, MA.
- MacLeod, W. B. och Miguel Urquiola. 2012. "Competition and Educational Productivity: Incentives Writ Large." IZA Discussion Paper No. 7063, IZA Institute of Labor Economics.

Malmberg, Bo, Eva Andersson och Zara Bergsten. 2013. "School Choice Motives: The Effects of Class and Residential Context." Working Paper 2013: 11, Stockholm University Linnaeus Center on Social Policy and Family Dynamics in Europe, Stockholm.

Markussen, Eifred, Martin Flatø och Rune B. Reiling. 2017. "Skolebidragsindikatorer i videregående skole." Rapport 2017:7, Nordic Institute for Studies in Innovation, Research and Education, Oslo.

Martinez, Erika. 2010. "Do Housing Prices Account for School Accountability?" Working Paper, Department of Economics, Duke University, Durham, NC.

Muriel, Alastair och Jeffrey Smith. 2011. "On Educational Performance Measures." *Fiscal Studies* 32(2):187–206.

Neal, Derek. 2010. "Aiming for Efficiency Rather than Proficiency." *Journal of Economic Perspectives* 24(3):119–132.

OECD. 2008. "Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools." Rapport, OECD Publishing, Paris.

Pope, Nolan. 2015. "The Effect of Teacher Ratings on Teacher Performance." Manuskript, University of Chicago.

Reback, Randall, Jonah Rockoff och Heather L. Schwartz. 2014. "Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind." *American Economic Journal: Economic Policy* 6(3):207–241.

Rockoff, Jonah E., Douglas O. Staiger, Thomas J. Kane och Eric S. Taylor. 2012. "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools." *American Economic Review* 102(7):3184–3213.

Rockoff, Jonah och Lesley J. Turner. 2010. "Short-Run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy* 2(4):119–147.

Rodas, Elizabeth I. R. 2015. "Do New Yorkers Vote With Their Wallets? The Impact of the Release of New York City Teacher Quality Data on Housing Prices and its Effects on Residential and School Segregation." Doktorsavhandling, Rutgers, The State University of New Jersey, Newark, NJ.

Rothstein, Jesse. 2017. "Measuring the Impacts of Teachers: Comment." *American Economic Review* 107(6):1656–1684.

Rouse, Cecilia E., Jane Hannaway, Dan Goldhaber och David Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." *American Economic Journal: Economic Policy* 5(2):251–281.

Ruijs, Nienke och Hessel Oosterbeek. 2017. "School Choice in Amsterdam: Which Schools are Chosen When School Choice is Free?" *Education Finance and Policy* https://doi.org/10.1162/edfp_a_00237:1–30.

Skolinspektionen. 2018. "Ombedömning av nationella prov 2017 – Fortsatt stora skillnader." Rapport, Stockholm.

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

Skolverket. 2017. "Yttrande över Samling för skolan. Nationell strategi för kunskap och likvärdighet (SOU 2017:35)." Dnr 2017:725, Stockholm.

- Skolverket. 2018a. "Skolutveckling." Stockholm. <https://www.skolverket.se/skolutveckling>.
- Skolverket. 2018b. "Välja skola." Stockholm. <http://www.valjaskola.se/>.
- Skolverket. 2018c. "Salsa – en statistisk modell." Stockholm. <https://www.skolverket.se/skolutveckling/statistik/om-skolverkets-statistik/salsa-statistisk-modell>.
- SOU 2017:35. "Samling för skolan – nationell strategi för kunskap och likvärdighet." Slutbetänkande av 2015 års skolkommision. Stockholm.
- Steffensen, Kjartan, Rachel Ekren, Oda O. Zachrisen och Lars J. Kirkebøen. 2017. "Er det forskjeller i skolers og kommuners bidrag til elevenes læring i grunnskolen? En kvantitativ studie." Rapport 2017/2, Statistisk sentralbyrå, Oslo.
- Stockholms stad. 2018. "Att mäta skolans bidrag till elevernas kunskapsutveckling." PM, Avdelning för utveckling och samordning/Uppföljningsenheten, Utbildningsförvaltningen.
- Strong, Michael, John Gargani och Özge Hacifazlıoğlu. 2011. "Do We Know a Successful Teacher When We See One? Experiments in the Identification of Effective Teachers." *Journal of Teacher Education* 62(4):367–382.
- Tåhlin, Michael 2007. "Överutbildningen i Sverige – utveckling och konsekvenser." Pp. 70–89 in *Utbildningsvägen – vart leder den? Om ungdomar, yrkesutbildning och försörjning*, redigerad av Johan Olofsson. Stockholm: SNS Förlag.
- Thomson, Dave. 2018. "Long term outcomes: Does school quality matter?" Blogg, Education Datalab, London. <https://ffteducationdatalab.org.uk/2018/12/long-term-outcomes-does-school-quality-matter/>.
- Totty, Evan. 2017. "High School Value-Added and College Outcomes." Manuskript, Department of Economics, Purdue University.
- Utbildningsdepartementet. 2012. "Uppdrag om skol- och resultatinformation." U2012/4307/S, Stockholm.
- Utdanningsdirektoratet. 2018. "Skolebidragsindikatorer for videregående skoler." Hemsida, Oslo. <https://www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/skolebidragsindikatorer-for-videregaende-skoler>.
- Walsh, Elias och Eric Isenberg. 2015. "How Does Value Added Compare to Student Growth Percentiles?" *Statistics and Public Policy* 2(1):1–13.
- Winters, Marcus A. och Joshua M. Cowen. 2012. "Grading New York: Accountability and Student Proficiency in America's Largest School District." *Educational Evaluation and Policy Analysis* 34(3):313–327.
- Winters, Marcus A., Julie R. Trivitt och Jay P. Greene. 2010. "The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam." *Economics of Education Review* 29(1):138–146.
- Witte, John F., Patrick J. Wolf, Joshua M. Cowen, Deven E. Carlson och David J. Fleming. 2014. "High-Stakes Choice: Achievement and Accountability in the Nation's Oldest Urban Voucher Program." *Educational Evaluation and Policy Analysis* 437–456.

www.svensktnaringsliv.se

Storgatan 19, 114 82 Stockholm

Telefon 08-553 430 00