

Bruk av informasjons- og belønningssystemer i offentlig skole

Hans Bonesrønning, Lars-Erik Borge, Torberg Falch og Bjarne Strøm

**ALLFORSK
2003**

ISBN 82-7570-145-7

FORORD

Denne rapporten om bruk av informasjons- og belønningssystemer i offentlig skole er utført på oppdrag fra Utdannings- og forskningsdepartementet og er organisert gjennom forskningsstiftelsen ALLFORSK ved NTNU. Prosjektet er utført av førsteamanuensis Hans Bonesrønning (prosjektleder), professor Lars-Erik Borge, førsteamanuensis Torberg Falch og førsteamanuensis Bjarne Strøm, alle ved Institutt for samfunnsøkonomi ved NTNU.

Trondheim, februar 2003

Hans Bonesrønning, Lars-Erik Borge, Torberg Falch og Bjarne Strøm

INNHold

side

Forord	iii
Innhold.....	v
1. Innledning og sammendrag	1
2. Dagens belønningssystemer	9
3. Kjennetegn ved et godt informasjonssystem.....	11
4. Utforming av informasjonssystemet	13
5. Innføring av informasjonssystemer – hva er erfaringene med indikatorene?	21
6. Kjennetegn og grunnlag for et godt belønningssystem.....	23
7. Hvem skal motta belønning?	25
8. Hvor stor andel av skolene bør motta belønning?	29
9. Vil insentivvirkningene avta over tid?	31
10. Virkemidler overfor skoler som fungerer dårlig	33
11. Innføring av belønningssystemer – hva er erfaringene?	35
12. Oppsummering og relevans for et norsk resultatbasert belønningssystem	45
Referanser	47

1. INNLEDNING OG SAMMENDRAG

Det primære formålet med etablering av informasjonssystemer i skolen er å gi elever, lærere, skoler og skoleeiere pålitelig informasjon om de resultater som oppnås, både absolutt sett og i forhold til andre. I mange tilfeller kombineres informasjonssystemet med et belønningssystem som gir belønning til aktører som oppnår gode prestasjoner, og eventuelt også sanksjoner mot de som oppnår utilfredsstillende resultater. Formålet med slike ordninger er å stimulere aktørene i sektoren til å arbeide mer beslutsomt for å oppnå gode resultater.

Per i dag har ikke Norge et utbygd informasjonssystem, men offentliggjøring av skolegjennomsnitt for elevenes standpunkt- og eksamenskarakterer ved avslutning av grunnskolen kan sees som et første steg i den retning. Vi har kimer til belønningssystem i form av bonusskoler, men bonustildeling er ikke avhengig av oppnådde resultater slik en vanligvis finner i belønningssystemer. Informasjons- og belønningssystemer er i rask utvikling andre steder. Kjernen i etablerte informasjonssystemer er indikatorer for kunnskapsnivå, og i noen grad faglig framgang, basert på elevresultater fra sentralt gitte kartleggingsprøver. Det store flertall av etablerte belønningssystemer er karakterisert ved at skoler belønnes eller sanksjoneres med basis i hvor store prosentandeler av elever som når definerte minimumsnivåer i lesing, skriving og/eller matematikk.

Informasjonssystemer kan i seg selv stimulere til gode resultater, selv om de ikke kombineres med eksplisitte pekuniære belønningssystemer. Det har for det første sammenheng med at informasjonssystemet vil ha elementer av ikke-pekuniær belønning og straff. For eksempel vil skoler som oppnår gode resultater få positiv oppmerksomhet i offentligheten, mens skoler som oppnår dårlige resultater kan oppfatte offentlig oppmerksomhet som stigmatiserende. Videre kan informasjonssystemet gi foreldrene et mer legitimt grunnlag for krav om endringer i organisering og drift av skoler som ikke oppnår tilfredsstillende resultater.

Informasjonssystemer kan for det andre bidra til å styrke de pekuniære insentivene i eksisterende finansieringssystemer. Dette poenget kan belyses gjennom to eksempler. La oss først tenke oss et system hvor skolene i betydelig grad finansieres gjennom skolepenger og hvor elevene fritt kan velge skole uavhengig av bosted. Innenfor en slik finansieringsmodell kan innføring av et informasjonssystem gi foreldre og elever et bedre beslutningsgrunnlag og på den måten bidra til mer effektiv konkurranse og bedre resultater. Tilsvarende mekanisme kan gjøre seg gjeldene innenfor et offentlig skolesystem med naboskoleregulering og hvor lokal eiendomsskatt utgjør en betydelig del av skolens finansiering. Poenget er at foreldre og elever da får et bedre beslutningsgrunnlag når de indirekte velger skole mellom valg av bosted. De pekuniære insentivene styrkes

fordi boligverdiene og dermed skolenes finansieringsgrunnlag vil være avhengig av de resultater som oppnås.

I dagens norske system kan det argumenteres for at de to mekanismene diskutert over er lite virksomme. Det har sammenheng med at elevene i begrenset grad kan velge skole uavhengig av bosted, at skoleeiernes (kommunenes) inntekter i liten grad påvirkes av det lokale eiendomsskattegrunnlaget, og at mobiliteten ofte antas å være lav. Dette kan være en begrunnelse for å kombinere informasjonssystemet med et belønningssystem. Et informasjonssystem som gir pekuniære insentiver gjennom et belønningssystem vil imidlertid være mer sårbart for utilsiktede virkninger enn informasjonssystemer som gir pekuniære insentiver på en mer indirekte måte gjennom andre finansieringsformer som brukerbetaling og lokal skattefinansiering. Det har sammenheng med at elever og foreldre kan basere sin valg av skole og bosted på et bredere og bedre informasjonsgrunnlag enn det indikatorene i et informasjonssystem vil kunne gi. Disse forholdene motiverer en grundig utredning av hvordan et informasjons- og belønningssystem bør utformes.

Formålet med dette notatet er å gi en diskusjon av hvordan et eventuelt norsk belønningssystem bør utformes. Først gis en diskusjon om hvordan informasjons- og belønningssystemet bør utformes med utgangspunkt i et sett av teoribaserte kriterier. Deretter belyses ved gjennomgang av empiriske analyser hvordan etablerte systemer påvirker atferden til elever, lærere, skoleledere og foreldre. Erfaringene gir innspill til hvordan en del av barnesykdommene ved slike systemer kan unngås.

Vi har satt opp følgende krav til et godt informasjonssystem: Det informerer om måloppnåelse på alle de viktigste områdene identifisert i måldokumentene, det skiller ut skolens bidrag til måloppnåelsen (dvs. tar vekk innflytelsen fra påvirkningsfaktorer utenfor skolen), informasjonen kan ikke manipuleres av aktørene, indikatorene uttrykker måloppnåelse og ikke tilfeldigheter som aktørene ikke har kontroll over, og informasjonen er lett forståelig.

Fem typer indikatorer er vurdert. Disse er indikatorer som gir gjennomsnittlig kunnskapsnivå i skolene, skolens bidrag til elevenes kunnskapsøkning i løpet av en periode, elevenes læringsstrategier, trivsel og foreldrenes brukertilfredshet. De to første typene er basert på nasjonale kartleggingsprøver, mens de tre siste er basert på spørreskjemaer til elever eller foreldre.

Indikatorene for elevenes kunnskapsnivå og læringsstrategier fanger opp viktige deler av skolens målsettinger, men langt fra alle målsettinger. De andre indikatorene er ikke like godt forankret i måldokumentene, men gir viktig informasjon om hvordan skolen oppfattes av brukerne. Indikatoren som gir elevenes kunnskapsøkning skiller seg ut fra de andre ved at den separerer ut skolens bidrag. De andre indikatorene reflekterer i tillegg til skolens bidrag også elevenes sosioøkonomiske bakgrunn og/eller evnenivå i større eller mindre grad. Alle indikatorene kan manipuleres av aktørene i skolen. Indikatoren for

kunnskapsøkning er vanskeligst å påvirke ved dysfunksjonell atferd, mens indikatorene basert på spørreskjemaer er enklest å manipulere av aktørene i sektoren. Alle indikatorer kan dekomponeres i signal og støy. Signalet formidler den ønskede informasjon som for eksempel den faktiske kunnskapsøkning i løpet av en periode, mens støy omfatter alt annet som at elevene har vært ekstra heldige med oppgavene, elevene har hatt en ekstra god dag, elevsammensetningen er mer gunstig enn hva vi er i stand til å fange opp og kontrollere for og liknende. Problemet med indikatoren for kunnskapsøkning er at signalet kan være svakt relativt til støy. Dette problemet er mindre for nivåindikatorer (kunnskapsnivå, trivselsnivå) hvor signalet er kraftigere.

Belønningsgrunnlaget i et eventuelt belønningssystem må hentes fra informasjonssystemet. På den ene siden er det ønskelig at belønningsgrunnlaget inneholder mange indikatorer for å unngå målforskyvninger, på den andre siden bør belønningsgrunnlaget ikke inkludere indikatorer som er lette å manipulere ved dysfunksjonell atferd fra aktørene i skolen. Det siste hensynet tilsier indikatorer basert på resultatene fra kartleggingsprøver, mens det første hensynet tilsier at en forsøker å etablere et bredere grunnlag som inkluderer læringsstrategier og trivsel. En mulig løsning vil kunne være å inkludere de sistnevnte indikatorene med noe mindre vekt enn indikatorene basert på kartleggingsprøvene.

Indikatorene basert på kartleggingsprøvene er av to typer og gir henholdsvis kvantifisering⁸ av gjennomsnittlig kunnskapsnivå og skolens bidrag til kunnskapsnivå. Valget mellom disse indikatorene vil være av avgjørende betydning for belønningssystemets funksjonsmåte. Bruk av kunnskapsnivå som belønningsgrunnlag har god intuitiv appell og signaliserer at en har store forventninger til alle typer elever. På den andre siden vil et slikt belønningsgrunnlag gi feilaktige insentiver i den grad belønningen tilfaller skoler med den mest gunstige elevsammensetningen selv om skoler med mindre gunstig elevsammensetning bidrar mer til elevenes kunnskapsøkning. Bruk av indikatorer som fanger opp skolens bidrag til kunnskapsøkningen gir riktigere insentiver, og vil lettere kunne oppfattes som rettferdige av aktørene i skolen.

Belønning bør gis til aktører som har muligheter til å påvirke kvaliteten. Vi ser her bort fra at belønning gis til elevene. Aktuelle kandidater er da enkeltlærere, skoler og skoleeiere - som alle disponerer virkemidler av ulike typer. Det er flere problemer med å belønne skoleeiere, men det viktigste er kanskje at de største eierne (for eksempel Oslo kommune) på grunn av 'store talls lov' i de fleste tilfeller vil plassere seg omtrent midt i utvalget og således sjelden motta belønning med mindre ordningene er svært omfattende. Problemene med å belønne enkeltlærere er at signalet vil være svakt relativt til støy pga få observasjoner, at enkeltlæreres bidrag kan være vanskelig å identifisere dersom for eksempel samarbeid praktiseres, og at kartleggingsprøvene må komme hyppig (dvs. årlig) og omfatte alle lærere slik at enkeltlæreres bidrag kan identifiseres og alle lærere ha muligheter til å motta belønning. Skolene kommer i en gunstig mellomstilling mellom enkeltlærere og skoleeiere.

I mange tilfeller vil imidlertid skolene bare ha en klasse på hvert årstrinn slik at det ikke vil være gevinster ved å beregne effekter på skolenivået framfor klassenivået. Skoleeffekten kan i slike tilfeller oppnås som et gjennomsnitt av prestasjonene over noen påfølgende år, mens det vil være vanskeligere å gjøre lærereffekten robust. Dersom en velger å benytte gjennomsnittet av prestasjonene over flere år skapes imidlertid noen nye problemer. Insentivene kan svekkes. Det vil for eksempel kunne gå flere år fra forbedringer av praksis faktisk skjer til belønning følger, fordi det er innbygde 'tregheter' i indikatorer som samler opp resultater over flere år. Hvor mange års gjennomsnitt som bør benyttes må framkomme ved en avveining mellom hensynet til robuste indikatorer og hensynet til insentivvirkningene.

Dersom en velger å belønne skolene, vil det være et spørsmål om belønningen skal gis av staten eller kommunene. Kommunene kan for eksempel avsette en pott til dette formålet og benytte informasjonen gitt i det nasjonale informasjonssystemet som grunnlag for tildeling. Alternativet er at staten gir belønning som kommer på toppen av den ordinære kommunale skolefinansieringen.

Generelt bør skoleledelsen få disponere belønningen fritt. Så lenge det foreligger insentiver til god ressursutnyttelse må en anta at skoleledelsen bruker pengene der kvalitetsforbedringen vil være størst – og det vil neppe være beslutningstakere plassert lenger opp i systemene som vil ha tilstrekkelig informasjon til å ta bedre beslutninger.

Hvor stor bør belønningen være? Når totalbudsjettet er gitt står en overfor en avveining mellom å gi relativt store beløp til noen få, eller relativt små beløp til mange. Det er flere grunner til at en bør velge den siste løsningen. For det første vil insentivene ha sterkest virkninger for skoler som ligger i nærheten av grensen for å få belønning, og med normalfordelte indikatorer vil antall skoler rundt den kritiske grensen øke når kravene for å oppnå belønning reduseres. For det andre trekker hensynet til mulig målforskyvning i retning av at belønningen ikke bør være for stor.

Et belønningssystem fungerer slik at skolene mottar belønning bare dersom prestasjonene gjentas for stadig nye grupper av elever. Dette tilsier at virkningene vil bestå over tid. Imidlertid kan en ikke se bort fra at det vil være mulig å kapitalisere historisk god skolekvalitet ved at den belønning som oppnås i de første rundene kan benyttes til å sikre suksess i senere runder. I så fall vil en kunne få et mer differensiert skolesystem hvor aktørene i de dårligste skolene gir opp. Selv om fremtidig suksess skulle være uavhengig av tidligere oppnådd belønning, vil noen skoler innse at sannsynligheten for å oppnå belønning er liten.

Det kan derfor være behov for spesielle tiltak for de lavest presterende skolene. I en del tilfeller vil en trusler om inngrep i form av restrukturering (som for eksempel utskifting av skoleledelse eller dårlige lærere), eller at elevene gis mulighet til å gå til andre og bedre skoler, kanskje være tilstrekkelig til å oppnå forbedring av resultatene. Dersom

forbedring ikke oppnås på denne måten vil det kunne være aktuelt å vurdere tiltak for restrukturering av skolene.

Det er særlige utfordringer knyttet til restrukturering av de dårligste skolene. For det første kan det være mange og ulike årsaker til dårlige resultater, og det vil kreves betydelig kompetanse for å stille riktig diagnose og identifisere de riktige tiltakene. Denne kompetansen må bygges opp - mest sannsynlig på statlig hånd. Staten må i så fall stille denne ekspertisen til rådighet for skoleeierne, i de fleste tilfeller vil det si kommunene. I den grad det vil være behov for tiltak bør dette skje gjennom samarbeidsavtaler mellom staten og skoleeieren. Det vil også kunne være aktuelt å vurdere om staten skal gå inn og gi elevene i de dårligste skolene rettigheter i form av fritt skolevalg.

Ytterligere innspill til hvordan informasjons- og belønningssystemer bør utformes kan fås ved å studere erfaringene gjort på steder der slike systemer har vært operative en tid. Vi har valgt å konsentrere oss om evalueringene gjort i fire amerikanske stater (Illinois (Chicago), Texas, Florida og North Carolina) fordi det her foreligger gode evalueringer.

Systemene etablert i disse statene er ulike langs de fleste dimensjoner, dvs. med hensyn til hvilke indikatorer som inkluderes, om vekten legges på belønning eller sanksjoner, størrelsen på eventuell belønning, om mye står på spill ('high stake') både for elever og skoler eller bare for en av partene. Evalueringene som er gjort så langt gir ikke grunnlag for å identifisere spesifikke elementer i systemene som gode og dårlige. Evalueringene handler mer om å identifisere noen grove typer atferdsrespons som følger uavhengig av spesifikke systemutforminger.

Erfaringene er at informasjons- og belønningssystemer gir betydelige atferdsendringer hos elever, lærere, skoleledere og foreldre. En entydig konklusjon er at elevprestasjonene – slik de måles ved kartleggingsprøver - forbedres signifikant over tid. Men det er også et nesten like entydig funn at elevprestasjonene – slik de måles ved kartleggingsprøver der lite står på spill ('low stake') – ikke forbedres like mye. (Det er en relativt opphetet diskusjon om innføring av informasjons- og belønningssystemer faktisk har bidratt til signifikante forbedringer i resultatene på kartleggingsprøvene der lite står på spill.)

Forskjellene i resultater mellom ulike typer kartleggingsprøver forklares med at lærere og elever tilpasser seg kartleggingsprøver der mye står på spill med å trene på de typer spørsmål som de vet disse prøvene inneholder, ved å bruke ekstra mye tid på forberedelser, og i noen tilfeller ved å anvende rent juks. Det er heller ikke uvanlig at lavt presterende elever holdes utenfor prøvene.

Foresattes respons på økt informasjon om skolekvalitet er forholdsvis sterk, og slår ut som økte boligpriser i områder med gode skoler. Ingen informasjonssystemer har eksistert lenge nok til at det er mulig å si noe sikkert om denne type respons fra foresatte er en nødvendig, eller tilstrekkelig, forutsetning for å generere atferdsendringer i skolene.

Belønningssystemer vil påvirke fordelingen av ressurser til skolene. Analyser gjort av systemet etablert i Florida viser at skoledistrikter med de største andelene av minoritets elever og elever fra lavinntektsfamilier vil kunne tape mest. Dette gjelder sannsynligvis ikke generelt, men reflekterer i Floridas tilfelle i stor grad belønningsgrunnlaget, som er andel elever som når definerte minimums kunnskapsnivåer.

I hvilken grad vil erfaringene gjort andre steder være relevante for Norge? Sannsynligvis er de meget relevante, men det er likevel grunn til å fremheve noen norske særegenheter. Kvalitetsportalen er planlagt med et mangfold av indikatorer, langt utover det en finner i etablerte informasjonssystemer. Virkningene av dette mangfoldet kan for eksempel være at indikatorer for trivsel og læringsstrategier fanger opp dysfunksjonell tilpasning til kartleggingsprøvene, og dermed medvirker til å begrense slik atferd. Mangfoldet i indikatorer åpner også for at foreldre kan vektlegge ulike egenskaper ved skolene, og at deres responser derfor kan gå i ulike retninger: Noen kan ønske seg til skoler der elevene trives godt, mens andre kan ønske seg til skoler hvor elevene lærer mye. I tillegg kommer at norske foreldres mobilitet er mindre enn amerikanske foreldres. Liten respons fra foreldre kan bidra til at informasjonssystemet i seg selv har små insentivvirkninger.

Hvordan kan vi utnytte erfaringene gjort andre steder til å forbedre utformingen av et eventuelt norsk informasjons- og belønningssystem? For det første identifiseres inflasjon i skårene på kartleggingsprøvene som et problem i et stort antall evalueringer. En grunn til slik inflasjon kan være at elevene drilles i kunnskaper og ferdigheter som er relevante for prøvene, men som ikke har mer generell anvendelse. Dersom dette er riktig, bør en forsøke å lage bedre prøver som det er vel verdt å undervise mot, dvs. en bør bestrebe seg på å prøve de ferdigheter og kunnskaper en ønsker å utvikle hos elevene. Kanskje vil det også være gevinster å hente ved å definere pensum skarpere slik at lærere og elever i større grad velger å forholde seg til pensum i stedet for til forrige års kartleggingsprøver.

En annen grunn til inflasjon i skårene kan være at elevene lærer å besvare oppgaver som gjenbrukes i kartleggingsprøvene. Dette er en mulig forklaring for resultatutviklingen i amerikanske kartleggingsprøver, som er karakterisert ved utstrakt gjenbruk av oppgaver fra et år til det neste. Denne praksis er motivert i ønsket om å fange opp utvikling over tid, men spørsmålet er om det vil være bedre å lage egne prøver der lite står på spill med formål å fange opp utviklingen over tid, og så redusere/ta bort gjenbruk av oppgaver i prøvene hvor mye står på spill.

Økt juksing er en tredje mulig forklaring på inflasjonen. Prøving av elever der mye står på spill (eksamen) er i de fleste systemer omgitt med tiltak som skal forebygge og avsløre forsøk på juks. Sannsynligvis vil det også være nødvendig å etablere ordninger som forebygger og/eller avslører juks ved kartleggingsprøver.

Den fjerde mulige forklaringen på inflasjon er at skolene i økende grad holder lavt presterende elever utenfor kartleggingsprøvene. Som tiltak mot denne utviklingen oppgir

mange amerikanske informasjonssystemer deltagelsesprosenter ved prøvene, og/eller hvor stor andel elever som er fritatt for deltagelse. Noen steder er stor prøvedeltagelse en nødvendig betingelse for å komme i betraktning for belønning. Det vil være viktig også for et norsk system å inkludere ordninger som stenger for at skolene kan oppnå gode resultater og belønning ved å ekskludere elever fra kartleggingsprøvene.

En styrke ved det amerikanske systemet er at det i lang tid har vært gjennomføres nasjonale kartleggingsprøver der lite står på spill. I dag kan disse prøvene benyttes som referansepunkt for evalueringer av de statlige informasjons- og belønningssystemene. Det bør vurderes om vi her i landet også bør etablere tilsvarende kartleggingsprøver. I tillegg til at disse prøvene vil være nyttige med hensyn til å vurdere informasjons- og belønningssystemenes virkemåte, vil de som nevnt ovenfor kunne brukes til å kartlegge norske elevers utvikling av kunnskaper og ferdigheter over tid.

Kan bonusskoleordningen eksistere videre i uendret form etter at informasjonssystemet er etablert? Dette vil være problematisk i den grad tildelingskriteriene i bonusskoleordningen viser seg å være forskjellige fra indikatorene i Kvalitetsportalen. En ville for eksempel kunne oppleve at noen bonusskoler presterer dårlig/har liten faglig framgang i ett eller flere fag, eller at trivselen i bonusskolen er dårligere enn i mange andre skoler. En slik situasjon vil være vanskelig å leve med. En kunne tenke seg flere løsninger. En mulighet er naturligvis å avvikle ordningen, en annen å endre tildelingskriteriene i bonusskoleordningen i tråd med indikatorene i Kvalitetsportalen, og la ordningen fortsette som starten på det nye belønningssystemet. En tredje mulighet kunne være å benytte ordningen til å stimulere eksperimenter med alternative undervisningsformer i skolene. Begrunnelsen for ordningen måtte da være at slik eksperimentering er viktig for å øke vår kunnskap om kjennetegn ved effektiv undervisning.

2. DAGENS BELØNNINGSSYSTEMER

De eksisterende finansieringssystemene for grunnskole og videregående skoler inneholder ingen *direkte* belønning knyttet til oppnådde resultater. Systemet er karakterisert ved at skoleinntektene følger elevtallet, at småskoler kompenseres for småskalaulemper, og at det gis ekstra tilskudd til noen elevgrupper (elever med behov for spesialundervisning, minoritets elever). For grunnskolen og store deler av videregående skole praktiseres naboskoleregelen, dvs. at elevene som regel må gå på skolen som er lokalisert nærmest hjemmet. Elevtallet varierer med naturlig demografisk variasjon, samt migrasjon inn og ut av skolenes definerte inntaksområde. Skolene opplever en sammenheng mellom oppnådde resultater/skolekvalitet og inntekter i den grad migrasjonen responderer på skolekvalitet. Vi vet lite om omfanget av skolekvalitetsmotivert migrasjon i Norge.

I enkelte fylkeskommuner praktiseres fritt skolevalg ved at elevene kan velge mellom (en delmengde av) fylkets videregående skoler, mens kapasiteten i skolene er fastlagt ved politiske beslutninger. Opptak til populære skoler rasjoneres ved karakterer. Det betyr at skoleinntektene ikke influeres av elevenes valg, men populære skoler vil ha elevgrupper med høy inntakspoengsum. Skolene gis dermed en ikke-pekuniær belønning for oppnådde resultater i den grad elevsøkningen reflekterer oppnådde resultater.

På toppen av de eksisterende finansieringssystemene er det i løpet av 2001-2002 etablert en bonusskoleordning for å belønne gode resultater og stimulere til lokalt utviklingsarbeid. Bonusskoler er skoler som etter søknad mottar 50 000 kroner fra UFD. Kriteriene for tildeling er knyttet til noen hovedområder: opplæring og læringsprosess, læringsutbytte og læringsresultater, læringsmiljø, utstyr og læremidler, ledelse og kompetanseutvikling og skoleanlegg.

I 2002/2003 fikk 92 skoler status som bonusskoler. En gjennomgang av begrunnelsene for tildeling viser at flertallet av de utvalgte skolene driver prosjekter innenfor feltene læringsmiljø og læringsprosesser. Det fremgår ikke at evaluering av læringsutbyttet er elementer i disse prosjektene. To av de utvalgte bonusskolene driver prosjekter direkte knyttet til læringsutbytte (evaluering av læringsutbytte/arbeid med læringsutbytte). Begge disse skolene inngår i forskningsprosjekter.

Bonusskoler kan oppfattes som et første steg i etablering av belønningssystemer for offentlige skoler. Nedenfor diskuteres om det er mulig å gå videre i denne retningen. Først gis en relativt grundig gjennomgang av den type informasjonssystem som må ligge til grunn for et eventuelt belønningssystem, deretter diskuteres ulike sider ved utforming av et belønningssystem. Generelt vil slike systemer ha både tilsiktede og utilsiktede virkninger. Utfordringen vil være å utforme systemet slik at en kan være rimelig sikker på å oppnå de tilsiktede virkningene, og slik at utilsiktede virkninger blir minst mulig.

3. KJENNETEGN VED ET GODT INFORMASJONSSYSTEM

I likhet med de aller fleste offentlig produserte tjenester er et hovedproblem i skolen å måle kvaliteten på de tjenester og den innsats som ytes og det resultat som frambringes. Den generelle teorien om forholdet mellom prinsipal og agent danner det teoretiske utgangspunktet for forståelsen av disse problemene¹. Prinsipalen (i vårt tilfelle skolemyndighetene) og agentene (skoleledere og lærere) forfølger ikke nødvendigvis de samme målsettinger. Siden det ikke er mulig for skolemyndighetene å observere innsatsen til de enkelte aktører, er utfordringen for skolemyndighetene å innhente informasjon om resultatene av aktiviteten og knytte belønning til disse resultatene for å gi agentene de riktige insentiver. Hvilke skoler er gode, hvilke er mindre gode, hvilke lærere er gode, hvilke er mindre gode? Dette er spørsmål som ikke bare er viktige for skolemyndighetene, men også for foreldrene når de skal velge bosted på grunn av naboskoleregelen. For å svare på slike spørsmål er det avgjørende at det foreligger informasjon om skolene. For å vurdere hva slags informasjon som skal innhentes og benyttes må det for det første foretas en avgrensing av de målsettinger skolen skal ha. Derneft må det hentes informasjon som belyser i hvilken grad disse målsettingene oppnås ved ulike skoler.

Målsettingene i norsk grunnskole; slik de er nedfelt i grunnskoleloven og læreplandokumenter kan grovt oppsummeres i to hovedelementer:

- A) Elevene skal tilegne seg kunnskaper og grunnleggende ferdigheter i lesing, skriving, matematikk og andre skolefag.
- B) Skolen skal bidra til at elevene utvikler gode læringsstrategier, sosiale ferdigheter, ansvarsfullhet, evne til å arbeide i fellesskap og være med på demokratiske beslutningsprosesser.

Målsettingene i skolen er altså svært mange, og selv om de ikke i utgangspunktet står i innbyrdes motstrid til hverandre, kan en gitt ressursramme innebære at tilgodeseelse av en målsetting vil føre til lavere oppmerksomhet om en annen. Kompleksiteten i målsettinger kan i utgangspunktet brukes til å stille spørsmål om det i det hele tatt er ønskelig og mulig å etablere et meningsfylt informasjonssystem for resultatene i skolen. I denne utredningen tar vi imidlertid som utgangspunkt at informasjon er ønskelig, selv om den ikke reflekterer situasjonen i forhold til alle målsettinger.

Vi har valgt å stille opp følgende kriterier som et godt informasjonssystem for skolen bør tilfredsstillere:

1. Gi informasjon om i hvilken grad de sentrale målsettinger for skolen nås
2. Identifisere skolens/lærernes bidrag til oppnåelse av målsettingene
3. Aktørene i skolen, lærerne og skolelederne, bør ikke ha muligheter for å manipulere informasjonen og informasjonsgrunnlaget

¹ En lettfattelig framstilling av den generelle teorien finnes i Gibbons (1998), mens Dixit (2000) gir en diskusjon av relevant teori for offentlig sektor og skolesystemet.

4. Gi så presis informasjon som mulig
5. Indikatorene må være lett forståelige

I tillegg bør informasjonssystemet som helhet ha lave kostnader.

Kriterium 1) følger direkte av den grunnleggende forutsetning at myndighetene har noen operasjonelle mål for aktiviteten i skolen. Dersom målsettingene for eksempel kan grupperes i de to elementene A og B beskrevet over, betyr det at systemet bør inneholde indikatorer som sier noe om i hvilken grad skolene oppfyller disse målsettingene. Dersom det bare inneholder indikatorer for B vil systemet på ulike vis kunne gi en uheldig målforskyvning.

Kriterium 2) innebærer at de indikatorer som brukes i informasjonssystemet gir informasjon om skolens og eventuelt lærernes bidrag til oppfyllelsen av målsettingene. Det betyr at en må kunne isolere effekten av skolens og lærernes effekt på indikatorene fra effektene av elevenes bakgrunn og evnenivå og den historisk innflytelse fra tidligere skolekarriere. Indikatorer som i overveiende grad avspeiler egenskapene ved elevenes hjemmemiljø, foreldrenes inntekt og utdanning vil være lite egnet som grunnlag for skolepolitiske beslutninger, og vil også kunne gi brukerne feilaktig informasjon om skolekvaliteten.

Kriterium 3) bygger på at prinsipalen (skolemyndighetene) og agentene (skolelederne og lærerne) kan ha ulike målsettinger og dermed kan agentene ha insentiver til å påvirke den informasjonen som gis til prinsipalen til å fremme egne målsettinger. Informasjonssystemet må innrettes slik at skolelederne og lærerne ikke kan manipulere informasjonsgrunnlaget slik at skolen framstår som mer vellykket enn den egentlig er. Et eksempel på denne type manipulering som er diskutert i litteraturen er at lavt presterende elever søkes unndratt kunnskapstester for å øke det gjennomsnittlige prestasjonsnivået ved skolen.

Utgangspunktet for kriterium 4) er for det første at verdien av informasjonen generelt øker med graden av presisjon i målingene. For det andre vil forskjeller i presisjon mellom ulike indikatorer i seg selv påvirke aktørenes atferd. For eksempel vil skolelederne ønske å konsentrere aktiviteten på områder der indikatorene er presise i forhold til områder hvor indikatorene er beheftet med mye støy og målefeil. Dette kan dermed gi uønsket målforskyvning.

Kriterium 5) tar som utgangspunkt at et informasjonssystem bør ha høy grad av legitimitet i potensielle brukergrupper; skolemyndigheter, foreldre, og andre. Det er derfor viktig at brukergruppene er i stand til å forstå hva indikatorene gir uttrykk for.

Endelig bør drifts- og etableringskostnadene må være på et nivå som gjør at de er lavere enn fordelene ved informasjonssystemet.

4. UTFORMING AV INFORMASJONSSYSTEMET

4.1 Indikatorer

I utgangspunkt må informasjonssystemet inneholde indikatorer som fanger opp elevenes kunnskaper, ferdigheter og holdninger – i tråd med læreplanenes målsettinger. Utover dette vil innholdet måtte avhenge av de formål informasjonssystem er tiltenkt. For eksempel vil det kunne være riktig å inkludere prosessindikatorer dersom formålet er å støtte skoler/lærere i videreutvikling av skolekvaliteten.

Vi har valgt å inkludere følgende indikatorer:

- Mål for kunnskapsøkning i basisfag
- Kunnskapsnivå i basisfag
- Læringsstrategier
- Trivsel
- Brukermedvirkning/Brukertilfredshet

Nedenfor gis en vurdering av disse fem indikatorene. Listen over indikatorer er ikke ment å være uttømmende, men vi tror den er lang og variert nok til å belyse de viktigste sidene ved ulike typer indikatorer. Det gis ingen vurdering av kostnadene ved de enkelte indikatorene, men i stedet en samlet vurdering av kostnadene ved informasjonssystemet i slutten av avsnittet.

A. Indikatorer på kunnskapstilegnelse

Vi starter med indikatorer som fanger opp elevenes kunnskapstilegnelse. Elevenes kunnskapstilegnelse påvirkes grovt sett av skolen og et sett påvirkningsfaktorer utenfor skolen. Dessuten er kunnskapstilegnelse kumulativ. Denne grunnleggende forståelsen motiverer inndelingen av indikatorer på dette området.

A.1. Mål for kunnskapsøkning i basisfag

Mål for kunnskapsøkning er en beregnet gjennomsnittlig klasse- eller skoleeffekt på kunnskapsøkningen i et fag for enkeltelever i løpet av en avgrenset tidsperiode. Etablering av slike mål skjer ved at elevene gis nasjonale kartleggingsprøver i de aktuelle fagene ved start og ved avslutning av den aktuelle perioden, og deretter at "differensen" mellom resultatene ved avsluttende og initial prøve etableres. Denne differensen er et mål på den samlede framgang i perioden og reflekterer både elevens kunnskaper, familie- og skoleinnflytelse. De beste målene for kunnskapsøkning oppnås ved å rense bort elev- og familieinnflytelse. Et rensset kunnskapsøkingsmål vil gi informasjon om klasse- eller

skolebidraget, men informasjonen har bare mening relativt til tilsvarende kunnskapsøkning for andre klasser eller skoler.

Styrken til mål for kunnskapsøkninger i første rekke at de identifiserer lærernes/skolenes bidrag til oppnåelse av skolens målsettinger. Målet har to egenskaper som sikrer dette. For det første er det rensset så langt som mulig for familieinnflytelse. I tillegg fanger det opp klassens/skolens bidrag for en nærmere definert tidsperiode. Dvs. en unngår at lærere på høyere trinn krediteres for arbeid som er nedlagt av lærere på lavere trinn.

Mål for kunnskapsøkning kan i prinsippet etableres for de fleste skolefagene, men vi antar her at bruk av kartleggingsprøver av praktiske hensyn og kostnadshensyn begrenses til basisfag som lesing/skriving (norsk), engelsk og matematikk. Det betyr at de kanskje mest sentrale fagene dekkes, men også at en rekke fag faller utenfor. En slik avgrensing kan medføre en målforskyvning ved at lærerne/skolen vrir mer ressurser fra fag uten prøver til fag med prøver.

Mål for kunnskapsøkning kan manipuleres ved at enkelte elevgrupper holdes utenfor prøvene, eller ved at elevene trenes hardt i relevante ferdigheter som anses relevante for prøvene på bekostning av mer grunnleggende begrepsforståelse. For å oppnå gevinster ved ekskludering av elever må skolene være i stand til å identifisere elever som oppnår mindre faglig framgang enn *tilsvarende* elever i andre skoler. Fordi det handler om å identifisere potensialet for relativ framgang, vil omfanget av slik manipulering sannsynligvis være beskjedent. Det er også grunn til å merke seg at skolene kan se seg tjent med relativt dårlige prestasjoner på den initiale testen slik at potensialet for framgang er større. Såkalte "teach-to-the-test" effekter er vanskelig å unngå. Dette potensielle problemet kan imidlertid snus til en styrke ved at testene lages slik at de fanger opp en betydelig del av fagets målsettinger. Det vil være viktig å kommunisere at fagplanene styrer utformingen av prøvene, og at dette gis troverdighet ved at skolene over tid erfarer en betydelig variasjon i oppgaver. Gjenbruk av oppgaver (som vil kunne være nødvendig for å fange opp utvikling over tid) må skje i relativt liten utstrekning og mønsteret bør ikke være lett gjennomskuelig.

Prøveresultater påvirkes alltid av tilfeldige faktorer: Noen elever kan ha hatt en dårlig dag, klassen er "heldig" med oppgavene, det er mye bråk i skoleomgivelsene osv.. Slike faktorer vil påvirke både initiale og avsluttende prøver, og bidra til målefeil/støy i "differensen". Dersom tidsavstanden mellom de to prøvene er liten – og den faglige framgangen følgelig liten – vil støyen kunne være stor i forhold til signalet, dvs. den faktiske faglige framgangen. Målet for kunnskapsøkning vil derfor kunne være beheftet med betydelig usikkerhet. Dette problemet kan motvirkes ved å sørge for en relativt stor avstand mellom prøvene, som bidrar til å forsterke signalet. Det vil imidlertid være en grense for hvor langt tidsspennet kan være. Først og fremst fordi formålet kan være å identifisere bidraget på ungdomstrinnet (3 år) eller i løpet av ett år i videregående skole. I tillegg kommer at optimal prøveavstand sannsynligvis vil avhenge av mobilitet i

elevflokken. Ved lav mobilitet kan en tillate større prøveavstand fordi en ikke risikerer at en stor andel av deltakerne ved den initiale prøven er borte ved den avsluttende prøven.

Det vil sannsynligvis ikke være enkelt å kommunisere informasjonen som ligger i indikatorene for kunnskapsøkning. En god forståelse forutsetter en forståelse av kunnskapsproduksjon som påvirket av både skoleinterne og skoleeksterne faktorer, samt en forståelse av kunnskapsoppbygging som en kumulativ prosess. Kommunikasjonsproblemene kan imidlertid snus til en fordel: Gjennom aktiv bruk av disse indikatorene kan en bidra til en mer opplyst diskusjon om skolekvalitet.

A.2. Kunnskapsnivå i basisfag

Kunnskapsnivået i et fag kan måles som gjennomsnittet av resultater fra kartleggingsprøver i faget, aggregert opp til klasse-, årstrinn- eller skolenivå. Gjennomsnittlig kunnskapsnivå reflekterer innflytelsen fra alle typer påvirkningsfaktorer, og er derfor ikke et godt mål på skolens bidrag.

Gjennomsnittlig kunnskapsnivå kan renses for innflytelse fra elev- og familiekarakteristika. En står da igjen med et mål som reflekterer skolens kumulative innflytelse på prestasjonsnivået. Det betyr for eksempel at et rensset gjennomsnittlig kunnskapsnivå ved utgang av 7. klasse vil kunne være et akseptabelt mål på barnetrinnets innflytelse (gitt at elevmobiliteten er beskjedent), mens tilsvarende indikator etter 10.klasse eller første år i videregående skole gir lite informasjon om ungdomstrinnets eller videregående skoles innflytelse.

Gjennomsnittlig kunnskapsnivå i en klasse eller skole kan relativt enkelt manipuleres av lærere og/eller skoleledelse ved å holde lavt presterende elever utenfor prøvene. Som for målene for kunnskapsøkning gjelder at skolene vil ha insentiver til "teach-to-the-test" atferd.

Måling av gjennomsnittlig kunnskapsnivå kan gjøres med de samme prøvene som benyttes til etablering av målene for kunnskapsøkning. Det betyr at målefeil/støy vil være den samme. I dette tilfellet vil imidlertid signalet være langt kraftigere, som betyr at problemet med signal i forhold til støy vil være lite alvorlig.

Indikatorer som gir gjennomsnittlig kunnskapsnivå i skolen er både lett forståelige og vanskelige å tolke. De gir lett forståelig informasjon om graden av måloppnåelse for elevene ved skolen. Hovedproblemet er at disse indikatorene ikke skiller mellom den gjennomsnittlige elevkvaliteten i skolen og skolekvaliteten. I noen tilfeller vil det være viktig å skille; spesielt mellom skoler hvor en har kombinasjoner av høy gjennomsnittlig elevkvalitet og dårlig skolekvalitet og skoler med lav gjennomsnittlig elevkvalitet og god skolekvalitet. Dersom indikatorene i første rekke reflekterer egenskaper ved elevene, og foreldre baserer sine skolevalg på disse indikatorene, vil begge de nevnte typene skoler

motta feil signaler fra markedet. Belønningssystemer som legger disse indikatorene til grunn vil kunne ha tilsvarende virkninger.

B. Andre indikatorer

Denne klassen av indikatorer avspeiler et sett av målsettinger for skolen som i utgangspunktet er videre enn det rent faglige og kunnskapsmessige innholdet.

B.1. Læringsstrategier

Læringsstrategier dreier seg om elevkompetanser på tvers av fag, og kan omfatte kontrollstrategier, som er elevenes evne til å overvåke egen læring eller utdypingsstrategier, som er elevenes evne til å sette nytt stoff inn i en sammenheng. En indikator for elevenes bruk av utdypingsstrategier kan etableres ved at elevene tar standpunkt til riktigheten av en rekke påstander av typen: "Når jeg arbeider med skolefag, forsøker jeg å knytte det nye stoffet til ting som jeg har lært i andre fag" og "Når jeg arbeider med skolefag, finner jeg ut hvordan stoffet passer inn i det jeg har lært tidligere". Prosedyren for etablering av en indikator for kontrollstrategier er tilsvarende.

Utvikling av gode læringsstrategier er en sentral målsetting i norsk skole, og det bidrar til en læringskapital som elevene kan utnytte i senere skolekarriere og arbeidsliv..

Indikatorene for læringsstrategier reflekterer innflytelse både fra skolen og en rekke skoleeksterne faktorer, og gir derfor liten informasjon om skolens bidrag. Indikatorene kan forbedres ved at de renses for familieinnflytelse. Det vil i prinsippet også være mulig å etablere mål som fanger opp skolens bidrag til læringsstrategier. (Det vil imidlertid være grenser hvor langt ned i klassene en kan gå med slike målinger, da spørsmålene som indikatorene er basert på krever et visst modenhetsnivå hos elevene).

Indikatorenes validitet avhenger av at elevene svarer ærlig på spørsmålene. Problemet er at bruk av indikatorene i et informasjonssystem kan legge press på elevene i retning av å overrapportere egne læringsstrategier. Dette presset vil kunne øke med den betydning indikatorene gis. De som måtte ønske å manipulere den informasjon som gis i spørreskjemaene må foreta mer subtile tilpasninger dersom en benytter indikatorer som fanger opp skolens bidrag til læringsstrategiene.

Indikatorer for læringsstrategier vil lett forståelig informasjon om måloppnåelse på disse feltene, men problemet er som for andre nivåindikatorer at skolens bidrag ikke kan separeres ut. Aktiv bruk av informasjonen kan gi uheldige insentivvirkninger som redegjort for under diskusjonen av indikatorene for kunnskapsnivå.

B.2. Trivsel

Trivselen til elevene kan måles ved spørreskjemaer til elevene hvor elevene blir bedt om å karakterisere egen trivsel direkte, eller ved at en trivselsindikator genereres med utgangspunkt i flere spørsmål til elevene. Trivselsnivået reflekterer summen av mange faktorer, som at elevene føler at de lærer mye (trekker sannsynligvis i positiv retning), og at de må anstrenge seg mye (trekker sannsynligvis i negativ retning). Mellommenneskelige relasjoner, som forholdet til medelever og forholdet til lærere, vil også påvirke trivsel i stor grad.

Læreplanene benytter ikke trivselsbegrepet direkte, men betydningen av et godt arbeidsmiljø understrekes. Kamp mot mobbing er en prioritert oppgave. Likeså understreker beslutningen om at Arbeidsmiljøloven skal gjelde for skolelever betydningen av at elevene trives på sin arbeidsplass.

Som for de fleste andre indikatorer vil en trivselsindikator avspeile innflytelse fra skolen og en rekke skoleeksterne forhold. Som for de andre indikatorene vil det være mulig å renske trivselsindikatoren for betydningen av familiebakgrunn.

Som andre indikatorer som er basert på spørreskjemaer til elevene vil validiteten være kritisk avhengig av at elevene svarer ærlig.

B.3. Brukervedvirkning/brukertilfredshet

Brukertilfredsheten til foreldre omfatter mange komponenter, som foreldrenes tilfredshet med skolens sosiale miljø og barnets trivsel, foreldermedvirkning og innflytelse, samarbeid skole og hjem, det pedagogiske tilbudet, tilpasset undervisning og elevenes fysiske arbeidsmiljø. Indikatorer for alle disse komponentene kan etableres ved spørreskjemaer til foreldrene.

Brukertilfredsheten til foreldre vil være et viktig signal til skolene. Spesielt vil svak brukertilfredshet være et varsel til skolen om at den står i fare for å miste - eller endre - rekruttering. Denne type informasjon muliggjør tidlige tiltak fra skolens side for å forbedre situasjonen.

Indikatorene vil reflektere en kombinasjon av faktiske/objektive sider ved skoletilbudet og foreldrenes forventninger. Alt annet likt vil brukertilfredsheten være lavest der forventningene er størst.

Foreldre kan ikke pålegges å delta i spørreundersøkelser. Det betyr at indikatorene neppe kan etableres på de laveste aggregeringsnivåene. Kanskje vil det være mulig å etablere indikatorene på skolenivå.

Foreldrenes insentiver til å feilrapportere sin tilfredshet med skoletilbudet avhenger av hvordan indikatorene vil bli brukt. Dersom brukertilfredshet inngår i et belønningssystem

vil foreldre kunne innse at overrapportering av brukertilfredshet bidrar til økte skoleinntekter kunne øke. Dersom brukertilfredsheten er inkludert i informasjonssystemet uten at det har direkte økonomiske konsekvenser for skolene, kan foreldre ønske å signalisere høye forventninger i håp om at dette ansporer skolene til større innsats.

4.2 Kostnader ved informasjonssystemet

Kostnadene ved informasjonssystemet er i første rekke knyttet til utvikling av kartleggingsprøver, til gjennomføring av prøvene, til vurdering av elevenes besvarelser, og rapportering av resultatene. Hoxby (2002) har beregnet disse kostnadene for en rekke amerikanske stater. Vi kan bruke Kentucky som eksempel. Denne staten har et ambisiøst system med indikatorer for kunnskapsøkning supplert med ekspertvurderinger av elevmapper. Omfanget av kartleggingsprøver er omfattende: Elevene prøves hvert år fra 3. til 10. klasse i forskjellige fag. Kostnaden ved evalueringen beregnes til omtrent 18 dollar per elev, som uansett hvordan en ser det er en lav kostnad.

Et eventuelt norske system innebærer sannsynligvis langt mindre prøving av elevene, og kanskje ingen mappeevalueringer. La oss se bort fra at disse forholdene trekker kostnadene ned sammenliknet med Kentucky. I stedet inkluderer vi kvalitetsforbedrende tiltak - som at prøvene gjøres mer raffinerte med lite gjenbruk av spørsmål, at kontroller i tilknytning til prøveavvikling etableres, at det vil være større kostnader enn i Kentucky med å utvikle indikatorene for kunnskapsøkning - og antar at denne type tiltak bidrar til fordobling av kostnadene. Kostnaden per elev vil da være rundt 300 kroner. Dette gir en indikasjon om at kostnadene ved informasjonssystemet vil være små.

4.3 Oppsummering

Tabell 1 gir en kompakt sammenligning av de fem indikatorene i forhold til kjennetegnene. I forhold til kjennetegn 1) om at indikatorene bør avspeile skolens målsettinger fungerer både kunnskapsøkning, kunnskapsnivå og læringsstrategi-indikatorene tilfredsstillende, mens trivsel og brukermedvirkningsindikatorene fungerer noe dårligere. Når det gjelder kjennetegn 2); muligheten til å isolere skolens bidrag scorer kunnskapsøkning høyt, mens alle de andre indikatorene er utilfredsstillende. Også når det gjelder manipulerbarhet (kjennetegn 3) kommer kunnskapsøkningens målet klart best ut. Vurdert etter kjennetegn 4 (signal/støy) har kunnskapsøkning klare svakheter og scorer lavest av indikatorene her, mens kunnskapsnivå her fungerer best. Også vurdert etter forståelighet er kunnskapsnivå den indikatoren som fungerer best.

Den foreløpige konklusjonen etter denne gjennomgangen er at kunnskapsøkning totalt sett kommer best ut av indikatorene. De andre indikatorene har alvorlige svakheter med å

isolere skolebidraget og/eller er betydelig mer utsatt for manipulering av aktørene i systemet. Samlet sett begrunner det at kunnskapsøkning bør utgjøre et hovedelement i informasjonssystemet. De andre indikatorene kan legges inn som et korrektiv begrunnet i hensynet til ivaretagelse av bredere målsettinger for skolen enn ren kunnskapstilegnelse.

Tabell 1. Sammenligning av indikatorene

	Målsetting	Skolens bidrag	Manipulerbarhet	Signal/støy	Forståelighet
Kunnskapsøkning	1	1	1	2	2
Kunnskapsnivå	1	3	2	1	1
Læringsstrategier	1	3	3	2	2
Trivsel	2	2	3	2	1
Brukermedvirkning	3	2	2	2	2

- 1: Meget bra
- 2: Middels
- 3: Problematiske

5. INNFØRING AV INFORMASJONSSYSTEMER – HVA ER ERFARINGENE MED INDIKATORENE?

Det gis en bred gjennomgang av erfaringene med informasjons- og belønningssystemer gis i avsnitt 12. Hovedvekten legges der på de atferdsendringer systemene genererer hos aktørene i og rundt skolen. Her gis en gjennomgang av erfaringene med indikatorer av kvalitet.

De fleste eksisterende informasjonssystemer har en kjerne av indikatorer basert på elevenes kunnskapsnivå i basisfagene. Bare et lite mindretall benytter mål for kunnskapsøkning som er generert fra gjentatt testing av de samme elevene over tid. Noen systemer lager imidlertid "kvasimål" for kunnskapsøkning ved å ta differensen mellom kunnskapsnivå i påfølgende kohorter i skolene. Andre indikatorer som inngår i noen systemer er drop-out rater, andel elever som deltar i testene, og elevenes tilstedeværelse i undervisning.

Evalueringsgjennomgangene gjort så langt identifiserer to hovedsvakheter ved indikatorer for kunnskapsproduksjon: De er ikke rene mål på skolebidraget, men reflekterer elevenes bakgrunn i større eller mindre grad, og de inneholder betydelig støy. Det er opplagt at gjennomsnittlig kunnskapsnivå reflekterer elevenes sosioøkonomiske bakgrunn, men Ladd og Walsh (2002) påviser at også indikatorer for kunnskapsøkning som benyttes i Nord- og Sør- Carolina er relativt sterk positiv korrelert med den sosioøkonomiske sammensetningen av skolens elevgrupper. Ladd og Walsh påpeker at mål for kunnskapsøkning som er influert av familiebakgrunn vil kunne ha uheldige konsekvenser ved at lærerne i økende grad søker seg bort fra skoler som underviser barn med mindre fordelaktig familiebakgrunn, men gir ikke empirisk belegg for at dette faktisk skjer.

Målene for kunnskapsøkning som er brukt i de to amerikanske delstatene er forskjellig fra målet foreslått for det norske informasjonssystemet. Det er derfor uklart i hvilken grad det foreslåtte målet rammes av denne kritikken, men uansett må det være en viktig del av kvalitetskontrollen å undersøke hvorvidt målet for kunnskapsøkning formidler systematisk skjev informasjon om skolens bidrag til elevenes kunnskapsutvikling.

Kane og Staiger (2001) gir en grundig analyse av støyproblemer i indikatorer for kunnskapsøkning. Støyproblemene skyldes to ulike forhold. For det første vil alle elevgrupper ha en del særegenheter (som ikke enkelt kan observeres eller måles). I små grupper vil variasjonen i prestasjoner som skyldes slike særegenheter ofte være stor relativt til den totale variasjonen i prestasjoner. Den del av variasjon som skyldes slike særegenheter fanges ikke enkelt opp av kontrollvariable (som foreldrenes utdanning og inntekt og liknende), men vil identifiseres som en del av skoleeffekten på prestasjonene. I store elevgrupper vil slike særegenheter ha mye mindre betydning. For det andre vil

prøveresultatene påvirkes av en del tilfeldige faktorer som ikke er følsomme for størrelsen på elevgruppen, slik som at det er mye forkjølelse/influenza i elevgruppen, at det er mye uro i omgivelsene på prøvetidspunktet, at det er uvanlig god "kjemi" i elevgruppen og liknende.

Selv små støyproblemer kan ha store konsekvenser. For det første fordi signalet i indikatorene for kunnskapsøkning (som er faktisk kunnskapsendring i løpet av en kortere eller lenger periode) kan være svakt, og for det andre fordi forskjellene i kunnskapsøkning mellom skoler i mange tilfeller vil være ganske liten.

Kane og Staiger dokumenterer at mindre enn halvparten av estimert kunnskapsøkning for lesing fra fjerde til femte skoleår for et utvalg elever fra Nord-Carolina skyldes varige forskjeller mellom skolene. Resten er støy. Løsningene de foreslår består grovt sett i å gjøre indikatorene mer presise ved å bruke mer informasjon, enten fra tidligere tester i samme fag eller fra tester i andre fag. Hvilken supplerende informasjon som bør brukes avhenger av korrelasjonene mellom ulike testresultater. For eksempel gir det god mening å etablere en indikator som gjennomsnittet av kunnskapsøkning over de siste årene dersom det viser seg at resultatene er relativt stabile over tid, men dette er en dårlig ide dersom resultatene endres mye fra et år til det neste. I slike tilfeller vil det kanskje være mer riktig å lage en indikator på tvers av fag.

Analysen til Kane og Staiger gir viktige innspill for det norske systemet. For det første indikerer den at det kan være riktig å basere seg på indikatorer for kunnskapsøkning som fanger opp kunnskapsendringen over flere enn ett år, for å sikre et relativt kraftig signal. For det andre gir analysen implikasjoner for hvilket nivå indikatorene bør etableres. Det at en del av støyen kan aggregeres bort, trekker i retning av indikatorer på høyere nivå enn klassen. For det tredje bør en vurdere tiltak for å gjøre indikatorene mer robuste ved å ta i bruk mer informasjon – når en har kommet så langt at slik informasjon foreligger.

6. KJENNETEGN OG GRUNNLAG FOR ET GODT BELØNNINGSSYSTEM

Et godt belønningssystem

- bygger på et godt informasjonssystem
- oppmuntrer aktørene til å bruke sin kunnskap produktivt, og oppmuntrer ikke til dysfunksjonell atferd
- belønner aktører som er i stand til å påvirke måloppfyllelsen
- berører en stor andel av aktørene
- sikrer vedvarende insentiver for kvalitets- og effektivitetsforbedring i skolen.

Belønningsgrunnlaget bør hentes fra informasjonssystemet, men må ikke nødvendigvis bestå av alle de inkluderte indikatorene. Introduksjon av belønning og sanksjoner gjør at aktørene i skolene får sterkere insentiver til å forbedre verdiene på indikatorene, både ved å forbedre elevprestasjoner – som er ønskelig – og ved ulike typer dysfunksjonell atferd – som oftest ikke er ønskelig. Det siste tilsier at indikatorer som er lette å manipulere bør holdes utenfor belønningsgrunnlaget, eventuelt tillegges liten vekt. På den andre siden vil introduksjon av belønning for et utvalg av indikatorene kunne bidra til målforskyvning ved at aktørene konsentrerer for mye ressurser om å oppnå gode resultater på de områdene som fanges opp av de inkluderte indikatorene på bekostning av de ikke-inkluderte. Det fins ingen enkle løsninger. Et godt belønningssystem vil være karakterisert ved en god balanse mellom de to hensynene.

Hensynet til dysfunksjonell atferd trekker i retning av å inkludere deltagerprosjenter ved prøvene i belønningsgrunnlaget.

Belønningssystemets virkemåte vil avhenge av om en velger indikatorene som gir kunnskapsnivå eller alternativt skolens bidrag til kunnskapsøkning. Bruk av kunnskapsnivå som belønningsgrunnlag har god intuitiv appell og signaliserer at en har store forventninger til alle typer elever. På den andre siden vil et slikt belønningsgrunnlag oppfattes som urettferdig, spesielt av skolene med den mest ugunstige sammensetningen av elevgruppen. Slike skoler kan gjøre en god jobb, men likevel oppleve at gjennomsnittsresultatene er dårligere enn i mange andre skoler. Bruk av indikatorene for kunnskapsøkning vil lettere kunne oppfattes som rettferdig av aktørene i skolen – i alle fall dersom det godtgjøres at de er uten sosioøkonomiske skjevheter. Ovenfor er det argumentert for at det vil være viktig å benytte indikatorer som er et gjennomsnitt for flere år slik at aktørene i skolene ikke oppfatter rangeringene som drevet av tilfeldige faktorer utenfor deres rekkevidde. Hvor mange års resultater som skal ligge til grunn må fremkomme ved en avveier et ønske om robusthet i indikatoren mot et ønske om at resultatforbedring i skolene skal fanges raskt opp i indikatorene.

7. HVEM SKAL MOTTA BELØNNING?

Ett av kjennetegnene ved et godt belønningssystem er at det belønner aktører som er i stand til å påvirke måloppfyllelsen. I forbindelse med etablering av et informasjons- og belønningssystem for grunnskolen er det tre naturlige kandidater; kommunen, den enkelte skole og den enkelte lærer.²

Kommunene treffer en rekke beslutninger som kan påvirke måloppnåelsen i skolene. For det første har kommunene et viktig finansieringsansvar siden de fatter vedtak om hvordan grunnskolen skal prioriteres i forhold til andre oppgaver. Kommunene påvirker også undervisningstilbudet gjennom fastsettelse av skolestruktur. En sentralisert skolestruktur med store skoler og mange elever per klasse gir et større faglig miljø for lærerne og mulighet for flere timer undervisning for elevene. På den andre siden vil små klasser og oversiktelige forhold kunne være en fordel ved en desentralisert skolestruktur. Videre vil kommunene kunne påvirke måloppnåelsen gjennom ansettelse av rektor ved den enkelte skole.

Den enkelte skole vil kunne påvirke måloppnåelsen på en mer direkte måte enn kommunen. Gitt at lærerne betyr mye for elevprestasjonene, vil skoleledelsen ha innflytelse på måloppnåelsen gjennom rekruttering av lærere og annet personell. Videre vil skoleledelsen gjennom rekruttering og andre tiltak kunne legge til rette for et godt arbeids- og fagmiljø blant lærerne. Endelig har skoleledelsen innflytelse på ressursdisponering, pedagogiske metoder og organisering.

Den utdanningsøkonomiske forskningen indikerer at lærerne betyr mye for elevenes faglige utvikling. Det har imidlertid vist seg vanskelig å påvise noen klar sammenheng mellom lærerkvalitet og lærerens utdanning og erfaring. Dette peker i retning av at det først og fremst er lærernes arbeidsinnsats, engasjement og undervisningsmetoder som har betydning for elevenes læringsutbytte.

Diskusjonen over viser at alle tre aktører (kommune, skole og lærer) disponerer virkemidler som kan påvirke måloppnåelsen. De avgjørende spørsmålene vil da være hvem som disponerer de viktigste virkemidlene og i hvilken grad det er mulig å identifisere den enkelte aktørs bidrag til måloppnåelsen.

Etter vår vurdering er det flere forhold som peker i retning av at det er bedre å gi belønning til den enkelte skole enn til den enkelte lærer. Et første moment er at et belønningssystem rettet mot den enkelte lærer måtte baseres på et snevrere utvalg av

² I prinsippet kan belønning for gode prestasjoner også gis til den enkelte elev. En diskusjon av slik belønning faller utenfor rammen av dette prosjektet som tar utgangspunkt i et informasjonssystem hvor skolens rolle står sentralt. Et slikt belønningssystem reiser også en rekke etiske problemstillinger.

indikatorer enn et belønningssystem på skolenivå. Det har sammenheng med at indikatorer som læringsstrategier, trivsel og brukertilfredshet vanskelig kan brytes ned på den enkelte lærer. Elevenes kunnskapsøkning i enkeltfag kan i prinsippet relateres til den enkelte lærer, men identifikasjon av lærereffekter vil være mer ressurskrevende enn identifikasjon av skoleeffekter. De nasjonale tester må i så fall dekke flere fag (alle lærere skal gis mulighet til belønning) og foretas hyppigere (skifte av lærer og lærermobilitet) for å identifisere lærerens bidrag til kunnskapsøkningen. Men selv om dette gjøres, kan det være vanskelig å identifisere den enkelte lærers bidrag, for eksempel fordi gode arbeidsrutiner som innarbeides i ett fag kan bidra til gode prestasjoner også i andre fag.

I denne sammenhengen er det viktig å presisere at belønningssystem på skolenivå praksis også kan innebære belønning på lærernivå. Dersom en skole mottar belønning, vil konsekvensene for den enkelte lærer avhenge av hvordan skoleledelsen disponere midlene. Midlene kan for eksempel anvendes til læremidler, lokaler og utstyr, til generell lønnsøkning til lærerne, eller til individuelle lønnstillegg. Dersom midlene i noen grad benyttes til individuelle lønnstillegg, blir skillet mellom belønningssystemer rettet mot henholdsvis skole og lærer mindre skarpt. Vi vil likevel hevde at belønning på skolenivå er å foretrekke siden skoleledelsen kan trekke på et bedre og bredere informasjonsgrunnlag enn indikatorene i informasjonssystemet når skolens beste lærer skal identifiseres.

Et belønningssystem på kommunenivå kan utformes på ulike måter. En mulighet er at belønningen baseres på gjennomsnittsverdiene for skolene i kommunen. Ulempen med et slikt opplegg er at det vil være vanskelig for store kommuner å oppnå belønning. På grunn av 'de store talls lov' vil store kommuner som regel komme ut rundt gjennomsnittet. Alternativt kan det nasjonale systemet baseres på informasjon på skolenivå slik at belønningen til den enkelte kommune avhenger av hvor mange 'gode' skoler det er i kommunen. Også dette alternative har svakheter. Det er ikke opplagt at en kommune med en 'god' og en 'dårlig' skole fortjener større belønning enn en kommune med to 'middel' skoler.

Videre er det viktig moment at de generelle argumentene for etablering av belønningssystemer er sterkere overfor de tjenesteproduserende enheter enn overfor kommunene. Det har sammenheng med at stat og kommune har sammenfallende interesser i den forstand at de ønsker et best mulig skoletilbud, mens lærere og skoleledelse i noen grad kan ha interesser som kommer i konflikt med dette (for eksempel knyttet til lønn og arbeidsinnsats). Dette tilsier at det er skolene, og ikke kommunen som bør motta belønningen.

Gitt at vi har argumentert for at skolene skal motta belønningen, kan det også være naturlig å spørre hvem som skal gi belønningen. Skal det være staten eller kommunene? Kommuner og fylkeskommuner har stor frihet når det gjelder utforming av finansieringsordninger for sine tjenesteproduserende enheter, men insentivbaser

finansieringsordninger er ikke etablert i særlig omfang. For enkelte tjenester kan dette begrunnes med at kommuner og fylkeskommuner ikke har evne til å bære den økonomiske risiko som insentivbaserte finansieringsordninger kan medføre. Innføring av innsatsstyrt finansiering av sykehusene ble blant annet begrunnet på denne måten. Det er vanskelig å se at dette argumentet har særlig relevans i vår sammenheng. Kommunene kan for eksempel sette av en begrenset pott som fordeles mellom skolene avhengig av deres prestasjoner.

Etablering av et informasjonssystem med omfattende bruk av kartleggingsprøver og brukerundersøkelser krever betydelige faglige og økonomiske ressurser. En mulig forklaring på at kommunene ikke har opprettet egne belønningssystemer kan være at det er et for stort løft for den enkelte kommune å etablere egne informasjonssystemer. Etablering av et nasjonalt informasjonssystem i statlig regi vil følgelig gjøre det lettere for kommunene å etablere egne belønningssystemer som enten premierer de beste skolene i kommunene eller skoler som hevder seg klart over landsgjennomsnittet. Lokale belønningssystemer kan supplere eller erstatte tilsvarende system på nasjonalt nivå.

8. HVOR STOR ANDEL AV SKOLENE BØR MOTTA BELØNNING?

Dersom totalbudsjettet til bonusordningen er gitt ("lukket" belønningsordning) må en foreta en avveiling mellom størrelsen på belønningen per enhet og antall enheter som mottar belønning. For å få fram noen av momentene i denne avveilingen, antar vi først som en forenkling at problemet er endimensjonalt i den forstand at det handler om å lage et system som får flest mulig aktører til å yte størst mulig innsats.

Et mulig alternativ er å kombinere høye krav for å få belønning kombinert med stor belønning for hver enhet. I dette tilfellet vil den store belønningen stimulere til stor innsats, men høy standard vil trekke i motsatt retning fordi mange enheter vil gi opp arbeidet med å møte standardene når kravene øker. Enheter med de beste lærerne vil imidlertid fortsette å anstrenge seg. Et mulig utfall er derfor at enheter med de beste lærerne anstrenger seg mest og mottar stor belønning, mens de øvrige enhetene ikke påvirkes av belønningssystemet.³

Alternativer som kombinerer lavere krav for belønning med liten belønning for hver enhet vil kunne gi mer gunstige utfall i den forstand at flere enheter vil øke sin anstrengelse en del. Optimumsløsningen er karakterisert ved at tapene fra redusert belønning per aktør oppveies av gevinstene ved at flere aktører anstrenger seg mer. Vi har ikke nok empirisk kunnskap til å kunne si hvor grensene bør gå.

Det er imidlertid flere argumenter som trekker i retning av at belønningen bør være relativt liten. Først og fremst målforskyvning: at den økte anstrengelsen kan gå på bekostning av fag eller aktiviteter som ikke er del av belønningsgrunnlaget. I tillegg er det rimelig å anta at de fleste indikatorer vil oppvise en tilnærmet normalfordeling. Det innebærer at ved kombinasjonen liten belønning/lave standarder vil mange enheter ligge i nærheten av grensen for belønning. Dette er gunstig fordi insentivvirkningene vil være størst for de enheter som ligger nærmest grensen.

Ut fra diskusjonen gitt foran er det rimelig å anta at skolen er den enhet som skal motta belønningen. En kan velge å gi likt beløp til alle skoler uavhengig av elevtall, eller en kan velge å la belønningen øke med elevtallet. Diskusjonen her tar utgangspunkt i det siste alternativet utformet som at skoler som er berettiget belønning mottar et fast beløp per elev.

Et belønningssystem med gitt totalbudsjett kan utformes på flere måter. En mulighet er først å fastlegge belønning per elev, og så fastsette standarden etter at resultatene foreligger slik at produktet av belønning per elev multiplisert med antall elever som er

³ Becker og Rosen (1992) gir en grundig teoretisk diskusjon av hvordan standarder og belønning vil påvirke elevatferd. Diskusjonen som gis her bygger på deres analyse.

belønningsberettiget er lik totalbudsjettet. En annen mulighet er å fastsette standarder først, og så tilpasse belønning per elev etter at resultatene foreligger slik at produktet er lik totalbudsjettet. Forskjellen på de to systemene er at skolene i det siste tilfellet kjenne kravene de må oppfylle for å oppnå belønning initialt, mens i det første tilfellet vil mulighetene for å oppnå støtte avhenge av atferden til de andre skolene i løpet av perioden som er til vurdering.

Dersom en ønsker at både standard og størrelsen på belønningen skal være kjent på forhånd vil totalutgiften være bestemt av hvor mange enheter som oppfyller standarder ("åpent" belønningssystem). Det er gode argumenter for en slik ordning, først og fremst knyttet til forutsigbarhet for aktørene i sektoren. Motargumentet er at en kan miste budsjettkontroll i makro, men sannsynligvis vil dette dreie seg om "småpenger" slik at hensynet til budsjettkontroll ikke vil veie tungt.

Valget av indikatorer i belønningsgrunnlaget vil legge føringer på hvilke alternativer en kan velge blant. Dersom en velger å benytte indikatorer for kunnskapsøkning vil det være mest nærliggende å oppgi antallet skoler (eller totalt antall elever dersom belønningen gis som kroner per elev) som skal motta belønningen, og så fastlegge standarden etter at evalueringsresultatene foreligger. Dette er fordi det er vanskelig å etablere absolutte standarder med utgangspunkt i indikatorer for kunnskapsøkning. Dvs. det kan være nærliggende å låse både antall enheter som får belønning og størrelsen på belønningen.

De fleste amerikanske belønningssystemer bygger på indikatorer for kunnskapsnivå. Systemene er karakterisert ved gitte standarder, for eksempel er standarden oppgitt som andelen av elevene som må prestere på eller over et gitt kunnskapsnivå, og alle som kvalifiserer mottar belønningen.

Størrelsen på belønningen varierer mye. I Texas har belønningssystemet beskjeden omfang, med lav belønning per elev og med mindre enn 10 prosent av skolene som kvalifiserer for belønning i 2000-2001. Floridas system er mye mer ambisiøst. Her gir 100 dollar per elev i skoler som er kvalifisert for belønning, og i 2000-2001 dreide det seg om 1300 av 3686 skoler (35 prosent).

9. VIL INSENTIVVIRKNINGENE AVTA OVER TID?

Kjernen i et belønningssystem må/vil være at prestasjonene må gjentas, dvs. skolene må skape gode resultater for stadig nye kull av elever for å motta belønning. Dersom belønning er basert på relativ rangering, må den enkelte deltaker i hver runde sørge for å prestere bedre enn de andre. Dette taler for at insentivene ikke vil svekkes over tid.

Men det er skjær i sjøen. For å se dette, ta utgangspunkt i at skolene har tre muligheter for å oppnå bedre resultater: Ulike typer dysfunksjonell atferd, forbedring av undervisningen til eksisterende lærerstab, og rekruttering av bedre lærere. Skoler som initialt klarer å oppnå gode resultater ved utstrakt dysfunksjonell atferd eller fordi de i utgangspunktet gir god undervisning vil ha en fordel ved at oppnådd belønning kan benyttes til å rekruttere gode lærere – som i neste runde bidrar til å gjenta suksessen. (Vi har her sett bort fra at indikatorene ikke er nøytrale i forhold til elevbakgrunn slik at skoler med gunstig elevsammensetning vil ha størst sannsynlighet for å motta belønning i de tidlige fasene.) Denne dynamikken trekker i retning av et todelt skolesystem med suksessrike skoler og mislykkede skoler – hvor den førstnevnte gruppen etter hvert vet at suksess er sikret nærmest uansett og hvor den sistnevnte gruppen resignerer.

Hvor sannsynlig er en slik likevekt?

Den avgjørende forutsetningen i resonnementet er at mer ressurser gir skolene et konkurransefortrinn som vil utnytte til å generere bedre resultater. Vi vet at det ikke er en enkel sammenheng mellom ressurser og resultater. Kanskje vil bedre insentiver generere en slik sammenheng mellom ressurser og resultater, men det er ikke opplagt.

Den langsiktige dynamikken som følger av et belønningssystem kan i stedet være at en mer resultatorientert skole vil virke attraktiv på de beste blant potensielle lærere slik at skoleledere vil oppleve at de har flere å velge blant ved ansettelser. I den teoretiske insentivlitteraturen fremholdes dette av og til som den viktigste effekt av et resultatbasert system. En utvidelse av poolen av gode lærere vil ikke bare komme de beste skolene til gode, men også mange av de mindre gode skolene. I det minste må mange gode lærere tilbringe år i mindre gode skoler mens de venter på ledige stillinger i gode skoler.

I tillegg kommer at informasjonssystemet gir muligheter for å lære av beste praksis. Informasjonssystemet kan utbygges slik at karakteristika ved beste praksis identifiseres (de store datamengdene må analyseres), og at denne informasjonen blir lett tilgjengelig for alle skoler og lærere. Dette vil sannsynligvis være en nødvendig forutsetning for at skolene skal kunne forbedres gjennom forbedret undervisning for eksisterende lærerstab.

Likevel kan det være grunner for å vie de lavt presterende skolene spesiell oppmerksomhet.

10. VIRKEMIDLER OVERFOR SKOLER SOM FUNGERER DÅRLIG

Et godt informasjonssystem vil ikke bare avdekke hvilke skoler som oppnår de beste resultatene, men også hvilke skoler som presterer dårlig. Det kan være vel så nyttig at et informasjonssystem avdekker hvilke skoler som fungerer dårlig som at systemet avdekker hvilke skoler som fungerer godt. Basert på en grunnideologi om at alle elever skal ha like muligheter, er det viktig å løfte prestasjonsnivået i skolene med dårligst prestasjoner. Håndteringen av slike skoler er derfor viktig. Skoler som fungerer dårlig kan være langt fra å oppnå en belønning i et belønningssystem, og derfor har et slikt system sannsynligvis liten effekt på disse skolene. I de aller fleste tilfeller der det er innført et belønningssystem er da også ulike virkemidler overfor skoler som presterer dårlig innbakt i systemet, se diskusjonen i avsnitt 11.

Hvilke virkemidler er mulig for å bedre prestasjonene i skoler som presterer dårlig? I utgangspunktet er det viktig at slike virkemidler ikke sees på som attraktive for skolene og kommunene, det kan føre til et ønske blant enkelte aktører om å prestere dårlig for å motta positive virkemidlene. Det impliserer at man bør være forsiktig med å innføre rene finansielle virkemidler overfor skoler og kommuner som presterer dårlig. Men det er mange andre alternativer, som kanskje kan kombineres med finansielle virkemidler.

For statlige virkemiddelbruk er det en klar begrensning i at det er kommunene som er skoleeiere. Det er ingen tradisjon for at staten kan diktere hva en kommune skal gjøre ved kommunale enkeltenheter som en enkelt skole så lenge virksomheten ligger innenfor lovens rammer. Prinsipielt har staten to muligheter hvis den ønsker å påvirke funksjonsmåten til enkeltskoler uten særskilte og direkte inngrep. For det første kan den inngå et samarbeid med den gjeldende kommune. Dette samarbeide kan inneholde klare planer om omstrukturering ved skoler som fungerer dårlig, for eksempel ved skifte av skoleledelse, assistanse fra ekspertteam, videreutdanning av lærere, omplassering av lærere eller endringer ved skolebygg. For å få etablert slike samarbeidsavtaler mellom stat og kommune er det mulig staten må dekke noe av kostnadene. Men ved etablering av samarbeidsavtaler mellom kommune og stat er det viktig at de utformes på en måte som gjør at de ikke framstår som attraktive for verken kommunene eller de ulike aktørene på skolene (skoleledelse og lærere). For eksempel kan videreutdanninga av lærere og opprusting av skolebygg være spesielt problematisk å inkludere i slik samarbeidsavtaler av denne grunn.

Behovet for samarbeidsavtaler avhenger av hva kommunene vil gjøre selv uten påvirkning fra staten. Et godt informasjonssystem vil avdekke for kommunene hvilke skoler de bør vie særskilt oppmerksomhet, noe de vil ha en egeninteresse i å gjøre uten at man trenger å trekke inn staten. Det er derfor mulig at et tilstrekkelig virkemiddel fra

staten sin side er å stille ekspertteam og lignende tilgjengelig når kommunene ønsker slik hjelp.

For det andre kan staten gi brukerne, det vil si elever og foreldre, nye rettigheter når det viser seg at de er knyttet til en skole som fungerer dårlig. Det kan for eksempel være en rettighet til å få velge fritt om man ønsker overflytting til en annen kommunal skole, eller en rettighet til overflytting til en privat skole uten å måtte betale skolepenger. Disse virkemidlene vil ha lavere kostnader enn samarbeidsavtaler, men kan ha liten effekt viss det er lang avstand til andre skoler.

Som utgangspunkt for en vurdering av hvilke skoler som fungerer dårlig kan det være hensiktsmessig å foreta en firedeling av skolene. En gruppe skoler vil kunne ha høy måloppnåelse og stor kunnskapsøkning. En annen gruppe vil kunne ha høy måloppnåelse og liten kunnskapsøkning (I dette tilfellet er måloppnåelsen generert av skoleeksterne forhold). En tredje gruppe vil kunne ha dårlig måloppnåelse og stor kunnskapsøkning, og en fjerde gruppe vil kunne ha dårlig måloppnåelse og liten kunnskapsøkning. Det vil være bred enighet om at skolene i den sistnevnte kategorien fungerer dårlig, og kanskje vil det være riktig å konsentrere innsatsen om denne gruppen. Det betyr i så fall at det bør ligge en vurdering av både prestasjonsnivået og kunnskapsøkningen til grunn for en beslutning om sanksjoner.

11. INNFØRING AV BELØNNINGSSYSTEMER – HVA ER ERFARINGENE?

I løpet av det siste tiåret er en rekke resultatbaserte belønningssystemer etablert i mange delstater i USA og i enkelte andre land. Antallet evalueringer basert på grundige empiriske analyser er foreløpig begrenset. De evaluerte belønningssystemene er forskjellige og ulike evalueringer har fokusert på ulike sider ved belønningssystemene. For eksempel legger flere evalueringer fokus på ulike typer dysfunksjonell atferd. Evalueringene er heller ikke av en slik art at det er mulig å isolere effekter av informasjonssystemet for seg fra effekter av at det er lagt belønning på toppen av informasjonssystemet. På denne bakgrunn er det foreløpig vanskelig å trekke sterke generelle konklusjoner om effekten av belønningssystemene. Vi vil likevel i det følgende forsøke å trekke ut av empirien hvilke effekter som ser ut til å være mest robuste.

Sammenligning av elevprestasjoner mellom stater med og uten resultatbaserte belønningssystemer på nasjonale prøver der lite står på spill er en måte å evaluere suksessen til belønningssystemene. Hanushek og Raymond (2002) benytter aggregerte tidsrekke-data for prestasjonene på de nasjonale prøvene i NAEP for ulike stater og finner at elevene presterer 1% bedre etter introduksjon av prøver der mye står på spill, men finner ingen signifikant forskjell mellom stater med og uten belønninger/sanksjoner knyttet til prøveresultatene⁴. De finner heller ingen klar sammenheng mellom endring i andelen spesialelever og introduksjon av prøver der mye står på spill når det tas hensyn til at denne andelen generelt har en trendmessig utvikling. Siden det bare er variasjoner mellom delstater som bidrar til identifisering av effektene i deres analyse, er det lite trolig at slike studier er egnet til å si særlig mye om de kausale sammenhengene mellom prestasjoner og "accountability". Et åpenbart problem er at innføring av "accountability" i en delstat ofte er et av flere elementer i en større reform av skolesystemet og dermed er det umulig å avgjøre hvorvidt de prestasjonseffekter og andre effekter som avdekkes skyldes det ene eller det andre elementet i reformene. Det er derfor av betydelig interesse å se på evalueringer basert på mer disaggregerte data.

De faglig mest omfattende og interessante evalueringene basert på disaggregerte data er gjort av systemene i Chicago, Florida, Nord-Carolina og Texas. For å lette oversikten og fordi systemene varierer i innretning og omfang, finner vi det hensiktsmessig å gå inn på erfaringene i hver av disse delstatene og deretter gir vi eksempler på analyser av mer spesifikke aspekter ved belønningssystemer basert på data fra andre delstater i USA samt Israel.

⁴ NAEP er forkortelse for National Assessment of Educational Progress.

Chicago:

Chicago innførte i 1996-97 et resultatsystem basert på kunnskapstesting. Systemet skiller seg en del fra mer konvensjonelle belønningssystemer ved at det inneholder insentiver både for lærere/skoleledere. Elever i 3., 6. og 8. klasse må tilfredsstille visse nasjonale minimumskrav i lesing og matematikk for å kunne flyttes opp i henholdsvis 4., 7. og 9. klasse. Videre gir ikke systemet direkte finansielle belønninger for de beste skolene, men retter oppmerksomheten mot de skolene som har størst problemer. Skoler hvor mindre enn 15% av elevene tilfredsstiller nasjonale normer i leseferdigheter settes under overvåkning med en trussel om at lærerne og administrasjonen ved disse skolene kan sies opp eller flyttes hvis ikke skolene forbedrer seg. Selv om systemet ikke er direkte sammenlignbart med det belønningssystemet som er skissert foran, er det av interesse å se på erfaringene fordi det belyser hvordan straffemekanismer for de dårligste skolene påvirker atferden og fordi det har vært gjenstand for grundige evalueringer av effektene [Jacob (2002), Jacob og Levitt (2002)]. Siden insentivene og selve testingen ble innført samtidig, er det ikke uten videre mulig å skille mellom den rene informasjonseffekten av testingen og den rene effekten knyttet til insentivene i systemet.

Evalueringene gir et klart inntrykk av at elevresultatene målt ved prøvene der mye står på spill forbedres kraftig etter innføring av insentivsystemet og at forbedringen er særlig høy hos de lavtpresterende skolene. Spørsmålet som deretter reises er hvorvidt dette reflekterer en økning i reelle faglige ferdigheter eller rene strategiske tilpasninger til systemet. For å undersøke dette sammenligner Jacob (2002) elevenes framgang på prøvene der mye står på spill med elevenes framgang på prøver ikke knyttet til insentivsystemet (prøver der lite står på spill) og i fag som ikke omfattes av prøvene der mye står på spill. Undersøkelsen viser at elevene ikke fikk tilsvarende økning i resultatene på prøvene der lite står på spill. Undersøkelsen viser spesielt at framgangen på matematikkprøvene der mye står på spill er sterk når det gjelder rene utregningsoppgaver og tallspørsmål, mens framgangen er liten på oppgavetyper som dreier seg mer om tolking og forståelse. Dette tolker forfatteren som indikasjoner på at lærerne systematisk har trenet elevene i egenskaper som lett kan påvirke prøveresultatene. Den manglende framgangen i fag som ikke inngår i kartleggingsprøvene (naturfag og samfunnsfag) tyder også på at lærerne responderer strategisk på innføring av insentivene.

Jacob finner også at andelen elever som definisjonsmessig ikke blir testet (for eksempel spesialelever) økte markant etter innføring av insentivsystemet, noe som kan skyldes at lærerne og skoleledelsen responderer på insentivsystemet med å systematisk klassifisere marginale elever som spesialelever. Videre finner de at skolene systematisk lar svake elever droppe ut eller gå om en klasse før de når det testede klassetrinn som respons på insentivsystemet.

Den mest dramatiske strategiske tilpasningen til insentivsystemet er ren juks med innrapporterte testresultater. Jacob og Levitt (2002a) utvikler en kontrollalgoritme for å avdekke mistenkelige trekk ved besvarelsene og mistenkelige sprang i testresultater i

enkeltklasser. Ved bruk av denne algoritmen anslås andelen av klassene i Chicago hvor alvorlig juksing foregår til å være 3-4% årlig. Videre tyder deres resultater på at omfanget av jukset reagerer sterkt på endringer i insentivene. Dette indikerer at ren juks kan være et alvorlig problem hvis ikke testingen følges opp av strenge kontrollrutiner. I en oppfølgingsstudie, Jacob og Levitt (2002b), hvor det brukes et utvalg av skoler som testes på nytt, viser de at den kontrollalgoritmen som de foreslår er svært effektiv til å avsløre juks med testresultatene.

Erfaringene fra Chicago identifiserer flere betydelige problemer med insentivsystemer basert på kunnskapstester. For det første er det klart at insentiver gir prestasjonsforbedringer på prøvene der mye står på spill, men at dette i betydelig grad skyldes rene strategiske tilpasninger til testene ("teach to test"). For det andre tilsier erfaringene derfra at prøver der mye står på spill må suppleres med prøver der lite står på spill for å kunne måle omfanget av utilsiktede strategiske tilpasninger til belønningssystemene.

Florida:

Florida startet en evaluering av sine skoler i skoleåret 1996/97 basert på kartleggingsprøver i matematikk og lesing. Initialt ble evalueringssystemet brukt til å identifisere lavtpresterende skoler, men i 1999 ble det laget en klassifisering av skolene på en skala fra A-F. Elevene i 4., 8. og 10.klasse blir testet i lesing og skriving, mens elevene testes i matematikk i 5., 8. og 10.klasse. Det er lagt inn belønninger og straffemekanismer i begge ender på dette evalueringssystemet. Skoler som får karakteren A, eller som forbedrer karakteren sin fra år til år mottar finansielle bonuser på ca. \$100 pr. elev, som kan brukes fritt. I skoler som får karakteren F gis elevene fritak for naboskoleregelen; elevene kan velge en annen offentlig skole med karakteren C eller bedre, eller de kan søke seg til en privat skole mens det offentlige betaler skolepengene. Slik sett er systemet i Florida en interessant konstruksjon med et resultatbasert belønningssystem med naboskoleregel som kjerne, og fritt skolevalg innebygd som straffemekanisme for de dårligste skolene.

En serie arbeider av David Figlio og flere medforfattere studerer ulike sider av Florida-systemet. Figlio og Lucas (2000) studerer hvorvidt den eksplisitte inndelingen i A-F skoler påvirker boligprisene når det samtidig kontrolleres for skår ved kartleggingsprøvene ved skolene. Tankegangen er at, gitt naboskoleregelen, vil graderingen bedre foreldrenes informasjon om skolekvalitet og føre til økt boligletterspørsel og økte boligpriser i områder med A-skoler. Deres resultater tyder på at selve graderingen i A-F-skoler gir disse effektene. De estimerte effektene innebærer at distinksjonen mellom A og B-skoler verdsettes til rundt \$9000 i boligmarkedet. Forfatterens tolkning er at denne effekten skyldes den ekstra informasjonen som inndelingen i A og B skoler representerer. Vi kan imidlertid ikke helt utelukke at foreldre tiltrekkes av skoler med mer ressurser, og at noe av effekten derfor skyldes at A-skoler belønnes eksplisitt med bonus.

Figlio og Getzler (2002) reiser spørsmålet om resultatsystemet i Florida har ført til manipulering av informasjonsgrunnlaget målt ved systematisk endring i andelen elever klassifisert som spesialelever. Deres empiriske undersøkelse er basert på et datasett for to fylker i Florida, der nesten 500 000 enkeltelever følges fra 1995 til 1999/2000, indikerer at introduksjon av prøver der mye står på spill innebærer 29% økning i andelen spesialelever. En særlig styrke ved deres undersøkelse er at de kontrollerer for alle elevspesifikke forhold som er konstant over tid. En hovedsvakheter er derimot at det bare er en kort periode etter innføring av systemet som studeres. Dersom klassifiseringsresponsen er av kortvarig karakter, vil deres resultater overvurdere de langsiktige problemene med denne type manipulering. Men studien gir i det minste klar indikasjon på at skolene initialt foretar strategisk klassifisering av elevene som respons på innføring av et testbasert evaluerings og belønningssystem.

Figlio og Page (2002) studerer de fordelingsmessige konsekvensene av å bruke en opsjon om fritt skolevalg for elevene som sanksjon mot de lavtpresterende skolene (F-skolene). Det gjøres ved å sammenligne de sosioøkonomiske karakteristika ved elevene på de naboskolene som gis fritt skolevalg basert på ratingen med hypotetisk skolevalgssystem der opsjonen om fritt skolevalg knyttes direkte til sosioøkonomiske karakteristika ved naboskolen. De finner at det alternative systemet med skolevalg direkte knyttet til sosioøkonomiske kriterier ved naboskolen vil fungere bedre fordelingsmessig enn Floridasystemet. Imidlertid gir ikke denne undersøkelsen noe svar på hvor effektivt fritt skolevalg som sanksjonsmiddel overfor de svakeste skolene er for å bedre prestasjonene sammenlignet med andre mulige tiltak overfor som overvåkning, utskifting av lærere og skoleledelse etc.

Texas:

Texas innførte i 1993 et testbasert evalueringsystem med innebygde insentiver knyttet til en eksplisitt "karaktersystem" for skolene og skoledistriktene. Studier basert på nasjonale prøver der lite står på spill som NAEP og kartleggingsprøven i TAAS der mye står på spill, referert i Carnoy, Loeb og Smith (2001) viser at skolene i delstaten hadde betydelig framgang mellom 1992 og 1996. Men studier basert på slike aggregerte data som NAEP er lite egnet til å avdekke reelle kausale sammenhenger mellom innføringen av dette systemet og elevenes kunnskapsnivå, mens framgangen på kartleggingsprøven TAAS der mye står på spill ikke uten videre kan tas som bevis på suksessen ved systemet på grunn av mulighetene for manipulering av elevgrunnlaget for prøvingen og "teach to test"-effekter beskrevet ovenfor. Studier basert på andre suksessindikatorer og mer disaggregerte data gir et mer nyansert bilde av effektene. I det følgende vil vi derfor gå gjennom resultatene i to av de mest metodisk solide studiene som er gjort. Testopplegget TAAS (Texas Assessment of Academic Skills) innebærer at elevene testes i lesing og matematikk i 3.-8. og 10. klasse og i skriving i 8. og 10.klasse. For å gå fra godkjent grunnutdanningen må de bestå TAAS-testen i 10. klassen. Dette er dermed en parallell til den nasjonale avgangseksamen i ungdomsskolen i Norge. Skolene deles inn i fire kategorier basert på aggregerte testresultater; "Exemplary", "Recognized", "Acceptable"

og "Low-performing". Kategoriene baseres på andelen elever som scorer over ett visst statlig fastsatt standardnivå på prøvene i lesing, skriving og matematikk. Utgangspunktet er altså kunnskapsnivå og ikke kunnskapsøkning. Videre kategoriseres elevene etter rase, og "economically disadvantaged" og andelen bestått må være nådd for alle kategorier elever. Videre må forrige års "dropout"-andel være under en viss grense, og andelen bestått må også vise en viss framgang fra forrige år. Hensikten er å rense bort innflytelsen fra sosioøkonomiske karakteristika og i størst mulig grad identifisere skolens bidrag til rangeringen. Rangeringen til de enkelte skoler offentliggjøres. Skoler med høyeste rangering fritas for en del reguleringer og inspeksjoner, og mottar også bonuser, men de finansielle belønningene er relativt ad hoc-preget. Skoler som faller i kategorien "lowperforming" blir gjenstand for spesiell evaluering og kan bli satt under administrasjon med ny skoleledelse hvis ikke resultatene forbedrer seg. Spørsmålet som reises i Carnoy, Loeb og Smith (2002) er hvorvidt suksessen i Texas målt ved TAAS-prøvene og NAEP kan gjenfinnes i andre, mer indirekte suksessmål for skolesystemet: Fullføringsandel for high school og tilbøyelighet til å ta høyere utdanning er to slike suksessmål. De finner at andelen av elevene som fullførte high school (dvs andelen 10.klassinger som fullførte 12.klasse) viste beskjedne økning i den perioden TAAS var implementert. Videre viser resultatene at andelen elever som planla høyere utdanning, (målt ved andelen "high school seniors" som tok standardiserte tester som er obligatorisk for opptak til høyere utdanningsinstitusjoner) økte i beskjedne grad. Dette indikerer at den kraftige resultatforbedringen målt ved andelen studenter som tilfredsstilte standardene i TAAS-prøvene på langt nær bekreftes av de andre suksessindikatorer. En må imidlertid være forsiktig med å tolke dette som et klart uttrykk for manglende suksess av evaluerings- og belønningssystemet. Det er fullt mulig at systemet har virket for kort tid til å ha fullt gjennomslag i disse alternative suksessindikatorer. Men det er likevel relevant å spørre om de sterke økningen i TAAS-resultatene i noen grad skyldes manipulering av elevgrunnlaget for prøvene.

For å rense informasjonen for støy og innflytelsen fra sosioøkonomiske karakteristika unntas en del elever fra evalueringsgrunnlaget på en skole et bestemt år: Tunge spesialelever, elever som ikke har engelsk som morsmål (i de tilfeller det ikke blir gitt spansktester), elever som er fraværende på prøvedagene og elever som gikk på vedkommende skole under et viss andel av skoleåret. I prinsippet har skolene en viss mulighet til å manipulere alle disse kategoriene. Cullen og Reback (2002) tar utgangspunkt i at insentivene til manipulering er sterkere for skoler som har resultater som ligger like under grensen for en bestemt rangering (for eksempel like under grensen mellom "exemplary" og "recognized") enn for skoler som ligger godt over grensen. De benytter data for testene og data for andelen unntak for skolene fra 1993 til 1998 for å undersøke om denne forskjellen i insentivene til manipulering slår ut i faktiske forskjeller mellom skolene i andelen elever som unntas fra å telle med i informasjonssystemet. De empiriske resultatene indikerer at dette er tilfelle. Undersøkelsen viser dermed at tiltak for å rense et nivåbasert informasjonssystem for støy og innflytelsen fra sosioøkonomiske

karakteristika kan gi økte muligheter for manipulering av systemet og identifiserer en vanskelig avveining for denne typen resultatbaserte evaluering og belønningssystemer.

Sett under ett er erfaringene fra Texas blandet. På den ene side er det dokumentert klare positive effekter på kartleggingsprøvene der mye står på spill, men på den andre siden tyder evalueringene på at de manglende effektene på andre suksessmål av denne forbedringen delvis skyldes dysfunksjonell atferd. Et interessant poeng ved Texas-systemet er at både skoler og skoledistrikter gis rating, men så langt har ingen forsøkt å sondre mellom effektene av rangeringen av skoler versus rangering av distrikter.

North Carolina:

North Carolina innførte et testbasert skolebelønningssystem (ABC) i skoleåret 1996-97. Innføringen kom delvis etter og delvis samtidig med en periode med betydelige reformer ellers i skolesystemet i delstaten, blant annet med økt lønn for lærerne og endringer i lærerutdanningen. Det er derfor vanskelig å fastslå om forbedringen i delstatens posisjon på de nasjonale testene (NAEP) i løpet av 90-tallet og relativ suksess også på andre resultatmål skyldes innføringen av belønningssystemet eller om det skyldes de andre reformene som ble gjennomført i samme tidsrom.

ABC-systemet bygger på leseprøver og matematikkprøver hvert år fra 3. til 8. klasse og skrivetester i 4. og 7. klasse. Det beregnes mål på kunnskapsøkning for hver skole. Det som legges til grunn for belønningen er skolens kunnskapsøkning målt i forhold til forventet standard. Beregningen av forventet standard varierer med fag og klasse, og delstatsgjennomsnitt, og mer tekniske detaljer om beregningen finnes i Ladd og Walsh (2002). Skolene klassifiseres i grupper etter i hvilken grad de tilfredsstill standarden. For eksempel klassifiseres skoler der forbedringen overstiger forventet standard med mer enn 10 % som "exemplary schools". Slike skoler mottar finansielle bonuser på \$1500 til lærere med godkjent utdanning (inklusive skoleledere) og \$500 til assistenter (support staff). På den andre siden vil skoler som havner i kategorien "low-performing" (mindre enn 50% av elevene tilfredsstill standarden) motta obligatorisk assistanse fra et delstatsteam og utsettes for sterk inngripen fra delstaten. Parallelt med innføringen av belønningssystemet ble skolelederne tildelt betydelig flere budsjettmessige og administrative virkemidler; blant annet allokering av budsjettet mellom undervisningspersonale og assistenter og unntak fra reguleringer av klassestørrelse og pensum.

Typen evalueringer som er foretatt er en spesiell grunn til å trekke fram erfaringene fra North Carolina. Evalueringen av insentivsystemene i Florida og Chicago som er referert over har vært basert på utviklingen i testresultater og andre målbare indikatorer på elevenes suksess. Disse undersøkelsene gir derfor ikke informasjon om hvilke og hvor stor del av de observerte effektene som skyldes endringer i skoleledelsens tilpasning, lærernes tilpasning eller elevenes og foreldrenes tilpasning. Ladd og Zelli (2001) gir derimot en eksplisitt analyse av hvordan belønningssystemet i North Carolina har endret

atferden til skolelederne. Dette er et viktig spørsmål, både fordi de fleste belønningssystemer har resultatforbedringer på den enkelte skole som målsetting, og fordi et resultatbasert belønningssystem må innebære at skoleledelsen får hånd om sentrale virkemidler som rekruttering av lærere, undervisningsmetoder og fordeling av lærere mellom elevgrupper og klassetrinn. Deres undersøkelse er basert på to runder med intervjuer av 70 tilfeldig utvalgte skoleledere i grunnskolen i delstaten og spørsmålene er utformet for å undersøke insentiveffekter basert på moderne prinsipal-agent-teori. Tre hovedkonklusjoner følger av deres undersøkelse: For det første har belønningssystemet uten tvil hatt betydelig effekt på skoleledernes atferd. For det andre har retningen på atferdsendringen i grove trekk vært i tråd med intensjonene i belønningssystemet som var å bedre elevenes ferdigheter i lesing, skriving og matematikk. Mange skoleledere responderte med å allokere mer ressurser til undervisningen i matematikk og lesing, og øke oppfølgingen av lærere i disse fagene. På den annen side illustrerer denne vridningen en av de viktigste problemene med målforskyvning i design av insentivsystemer i institusjoner med mange målsettinger og varierende muligheter for å verifisere resultatene i de ulike dimensjonene. Siden Ladd og Zelli fant at atferden til skolelederne ble betydelig endret er det et argument for å holde belønningen knyttet til de målte resultatene på et lavt nivå og utvide belønningsgrunnlaget som i dag består av testresultatene i basisfagene med andre indikatorer.

Den tredje konklusjonen gjelder fordelingsaspektet. Undersøkelsen indikerer at skolelederne faktisk økte oppmerksomheten om de svake elevene, hvilket er i tråd med intensjonene i belønningssystemet. Men hvorvidt de svake elevene totalt sett kommer bedre ut avhenger av hvordan belønningssystemet påvirker allokeringen av skoleledere og lærere mellom skoler. Dette gir ikke undersøkelsen noe direkte svar på, men forfatterne antyder at lærere og skoleledere i lavtpresterende skoler responderer på insentivsystemet med å flytte til skoler som mottar finansielle bonuser. Dette er potensielt viktige effekter, og nettoeffekten på de ressurssvake elevene kan derfor fort bli negativ. Siden Ladd og Zelli baserer seg på subjektiv informasjon gitt av et lite antall skoleledere er det selvsagt begrenset verdi av denne undersøkelsen. Men styrken og retningen på de effektene de avdekker er likevel så tydelige at de gir klare indikasjoner på at skolelederne reagerer som forventet på resultatbaserte belønningssystemer.

Andre undersøkelser:

Effekt på lærerallokering og læreratferd

De undersøkelsene som er referert over er alle (bortsett fra Ladd og Zelli (2001)) opptatt av å måle effektene av testbaserte informasjons- og belønningssystemer på prestasjoner (eller andre outputmål) og dysfunksjonell atferd. Men siden lærerne er sentrale aktører i kunnskapsproduksjonen er det rimelig å tro at eksplisitt belønning og straff til skolene knyttet til elevprestasjoner først og fremst virker gjennom læreratferd, lærerallokering og rekruttering. Dysfunksjonell atferd av samme karakter som manipulering av elevklassifiseringer kan forekomme dersom skolen systematisk allokterer de beste lærerne til de klassetrinn og de fag som testes. Det foreligger dessverre få grundige studier av dette

foreløpig. Boyd et al. (2002) studerer hvordan innføring av et offentlig tilgjengelig testbasert informasjonssystem for skolen i staten New York fra 1998-99 har påvirket lærerallokeringen på skolene. Staten New York tester elevene i fagene engelsk og matematikk i 4. og 8.klasse. I utgangspunktet har derfor skolene insentiver til å allokere de beste lærerne til disse fagene og disse klassetrinn. Samtidig kan insentivsystemet påvirke lærernes egne valg; for eksempel kan noen lærere ønske å forlate disse klassetrinn for å unngå den ekstra oppmerksomheten som det vil innebære å undervise på et klassetrinn som er gjenstand for testing. En mulig empirisk prediksjon som følger av dette er høyere utskifting av lærere på de klassetrinn som testes. Men det vil være vanskelig å avgjøre hvorvidt en eventuell effekt på lærerutskifting skyldes lærernes egne valg, eller bevisst omallokeringspolitikk fra skoleledelsens side. Boyd et al. (2002) benytter data for alle lærere og alle skoler i staten New York og finner overraskende nok at turnover for lærere i 4.klasse faktisk ble redusert umiddelbart før og etter innføringen av prøving der mye står på spill. Denne konklusjonen er robust over alle typer skoler fordelt etter prestasjonsnivå, sosioøkonomiske karakteristika og regioner. Men samtidig finner de en viss tendens til at de lærerne som tross alt sluttet i 4.klasse systematisk ble erstattet av lærere med lenger erfaring. Disse resultatene gir imidlertid begrenset informasjon om betydningen av strategisk lærerallokering. Siden kunnskapsproduksjon akkumuleres over tid kan det være rasjonelt for skoleledelsen å allokere de beste lærerne på klassetrinn forut for det trinn som er gjenstand for testing; i dette tilfellet 1.-3.klasse.

Lavy (2002a)* og Lavy (2002b) gir eksplisitte analyser av belønningssystemer og lærereffekter i Israel. I Israel ble 62 ungdomsskoler ("secondary schools") i 1995 plukket ut til å delta i et eksperiment der en bonus; delvis øremerket til flatt lønnstillegg til alle lærerne på skolen og delvis øremerket generelle forbedringer i arbeidsmiljøet for lærerne, ble gjort avhengig av skolens relative prestasjonsforbedring. Dette er altså i prinsippet et resultatbasert belønningssystem der belønningen er eksplisitt øremerket lærerne. Samtidig gjennomførte myndighetene et forsøksprogram der 22 high-schools (av en gruppe på 75 søkere) ble tildelt ekstra ressurser i form av økt undervisningstid og ressurser til opplæring av lærerne. Dette programmet kan betraktes som et mer tradisjonelt ressursvirkemiddel overfor skolene. Lavy (2002a) studerer effekten av disse to virkemidlene på ulike suksessindikatorer; elevprestasjonene, tilbøyelighet til å ta videre utdanning drop-out-tilbøyelighet og foretar en sammenligning av kostnadseffektiviteten av virkemidlene. Ved å bruke øvrige tilsvarende skoler i Israel som kontrollgruppe og benytte moderne økonometriske metoder til å fjerne seleksjonsskjevheter i "behandlingsgruppene" finner han betydelige positive effekter av begge virkemidlene på alle de undersøkte suksessindikatorene. Men det mest interessante funnet er at målt ved suksessøkning i forhold til utgifter, var ressurstildeling basert på resultater klart bedre enn den tradisjonelle metoden. Så langt er dette en sterk indikasjon på at resultatbaserte belønningssystemer er effektive virkemidler til å øke prestasjonene. Siden evalueringen bare omfatter et belønningssystem der bonusen er øremerket lærerne er det imidlertid vanskelig å si om et slikt øremerket belønningssystem er bedre eller dårligere enn et

belønningssystem der bonusen tildeles skolen som frie midler. Framtidig empirisk forskning kan forhåpentligvis gi svar på dette spørsmålet.

Et relevant spørsmål i forbindelse med lønnsbonuser for lærerne er hvorvidt slike bonuser bør tildeles enkeltlærere eller lærerkollegiet som helhet med flate tillegg, slik det ble gjort i eksperimentet diskutert ovenfor. Lavy (2002b) studerer effekten av et alternativt eksperiment i Israel der enkeltlærerne på "high schools" ble belønnet etter hvor godt deres elever forbedret eksamensresultatene i engelsk, hebraisk og matematikk, altså et rent prestasjonslønnsystem for enkeltlærerne. En hovedkonklusjon er at dette prestasjonslønnsystemet ga betydelig resultatforbedring i de fagene som inngikk i belønningsgrunnlaget, samtidig som det ikke hadde negative effekter på prestasjonene i de andre fagene. Han sammenligner dette individuelle prestasjonslønnsystemet med systemet med bonus for lærergruppen som helhet og finner at målt ved prestasjonsforbedring i forhold til utgifter er det individuelle systemet mer effektivt enn gruppebonus-systemet. Men den generelle gyldigheten av denne konklusjonen er mer usikker siden sammenligningen er basert på to forskjellige eksperimenter foretatt på ulike tidspunkter og under ulike rammevilkår.

12. OPPSUMMERING OG RELEVANS FOR ET NORSK RESULTATBASERT BELØNNINGSSYSTEM

Gjennomgangen ovenfor dokumenterer at det er sterkt økende forskningsinteresse rundt effekten av resultatbaserte belønnings og informasjonssystemer. Men når de empiriske resultatene skal tolkes må en ta i betraktning at systemene har virket i relativt kort tid, og det er dermed vanskelig å vite i hvor stor grad de effekter som avdekkes er av midlertidig eller varig karakter.

For det første er det et generelt bilde at elevprestasjonene målt ved kartleggingsprøver der mye står på spill viser klare forbedringer etter innføring av belønningssystemene. Men samtidig er det relativt klart dokumentert at en del av den initiale forbedringen representerer en slags "inflasjonseffekt" som dels skyldes at lærerne trener elevene bevisst i prøvefagene og i å besvare spørsmål i prøvene ("teach to test") og dels at elevene selv blir mer fortrolig med prøvesituasjonen. I hvor stor grad denne "teach to test"-effekten er uheldig kommer an på hva som er målsettingen for myndighetene. Dersom hensikten nettopp er å øke ferdighetene på de områder som prøves er det en ønskelig konsekvens av reformen. Dersom hensikten derimot er å nå et bredere spekter av målsettinger kan det virke uheldig. Men de studier som foreligger er ikke godt egnet til å svare på i hvor stor grad andre ferdigheter og målsettinger påvirkes negativt av belønningen og informasjonssøkningen om skolens resultater knyttet til de som faktisk testes. For øvrig er "teach to test"-mekanismer ikke ukjente fenomener i norsk skole. Trening i løsning av eksamensoppgaver som er en etablert og fullt akseptert praksis i norsk skole er et åpenbart eksempel i så måte. Det nye er eventuelt at insentivene til dette blir sterkere både for lærerne og skolen som helhet. Generelt tilsier de erfaringene som er høstet i andre land at belønningene knyttet til informasjonssystemet bør være relativt små, i hvert fall i en overgangsperiode før man har fått egne erfaringer om hvor godt systemet virker.

Et annet hovedresultat er at insentivene til manipulering av grunnlaget for de resultatbaserte belønningssystemene er sterke og at de faktisk manifesterer seg i konkrete handlinger fra aktørenes side. Det dreier som om alt fra systematisk manipulering av hvilke elever som danner grunnlag for de målte testresultatene til ren juksing med innrapporterte resultater. Det innebærer at den detaljerte utformingen av det informasjonssystemet som ligger i bunnen blir kritisk for hvor godt belønningssystemet vil virke. Ta manipulering av hvilke elever som klassifiseres som spesialelever og dermed unntas fra prøvingen som eksempel. I et system hvor prestasjonsnivå brukes som indikator vil det være insentiver til å la marginale elever klassifiseres som spesialelever og dermed unntas fra prøving. En mulig respons er å definere egne og "lettere" prøver for spesialelevne (eller generelt generere ulike krav til ulike studentgrupper for å fange opp sosioøkonomiske forhold). Men insentivene til å klassifisere en "vanlig, men marginal

elev" til denne gruppen vil åpenbart fortsatt eksistere, slik at dette ikke er noen løsning. Erfaringene fra Texas som har et slikt system bekrefter dette. Det er derimot klart at mulighetene og kostnadene ved manipulering er betydelig mindre når identiske elevers kunnskapsendring ligger til grunn for informasjonssystemet (kunnskapsøkingsmål). For at det på et tidspunkt skal lønne seg å klassifisere en elev som spesialelev må det innebære at man har en rimelig sikker prediksjon om *kunnskapsøkningen* til denne eleven relativt til andre elever flere år fram i tid. Grunnlaget for å foreta en vellykket manipulering på denne måten vil være betydelig mindre enn i et nivåbasert system. Dette er et tungt argument for å legge et system basert på indikatorer for kunnskapsøkning til grunn for måling av faglige ferdigheter. Samtidig er også de dokumenterte manipuleringene i de amerikanske systemene et sterkt argument for å la indikatorer som trivselsmål og subjektive rapporteringer om læringsstrategier ha relativt lav vekt i et samlet resultatmål. Grunnen er ganske enkelt at slike indikatorer åpenbart lar seg lett og relativt kostnadsfritt manipulere.

Videre er det svært viktig at kartleggingsprøver der mye står på spill suppleres med nasjonale prøver der lite står på spill for å avdekke størrelsen på den rene inflasjonseffekten av prøvesystemet og for å avdekke ren juks og manipulering. I det norske systemet kan eksamenskarakterene i en viss grad ha en slik kontrollfunksjon overfor prøveresultater på gitte tidspunkter. Men for at eksamenskarakterene skal ha denne funksjonen kan ikke gjennomsnittlige eksamenskarakterer ved hver skole være offentlig tilgjengelig informasjon, slik tilfellet er i dag. Det enkle poenget er at hvis denne informasjonen gjøres tilgjengelig vil eksamenskarakterene fungere som en kartleggingsprøve der mye står på spill for skolen (og ikke bare elevene). En annen begrensning ved eksamenskarakterene som kartleggingsprøve der lite står på spill for skolene er at utvikling i for eksempel eksamenskarakterer i matematikk i ungdomsskolen over tid trolig gir lite eller ingen informasjon om utvikling av kunnskapsnivået over tid. Et alternativt opplegg for å fange opp kunnskapsutviklingen over tid er å supplere kartleggingsprøver der mye står på spill med kartleggingsprøver der lite står på spill med relativt like oppgaver fra år til år. Et relevant praktisk spørsmål i denne forbindelse er hvorvidt disse prøvene hvor lite står på spill skal gis til alle skolene eller bare til et mindre tilfeldig uttrukket utvalg. Informasjonen kan selvsagt bli mer presis dersom alle deltar, men samtidig kan antallet tester samlet sett fortone seg som overveldende, og føre til at den enkelte skole og elev legger liten vekt på disse testene. Også rene kostnadshensyn tilsier at denne type prøver der lite står på spill foretas på et mindre utvalg av skoler.

REFERANSER

- Becker, W.E. og S. Rosen (1992): The learning effect of assessment and evaluation in high school. *Economics of Education Review*, 11, 107-118
- Boyd, D., H. Lankford, S. Loeb og J. Wyckoff (2002): Do high-stakes tests affect teachers' exit and transfer decisions? The case of the 4th grade test in New York State. Mimeo Stanford University
- Carnoy, M., S. Loeb og T. L. Smith (2001): Do higher state test scores in Texas make for better high school outcomes? Consortium for Policy Research in Education, University of Pennsylvania, Graduate School of Education, Report Series 047.
- Dixit, A. (2002): Incentives and organizations in the public sector. An interpretative view. *Journal of Human Resources* 37, 696-727
- Figlio, D. N. (2002): Fiscal implications of school accountability initiatives. NBER Tax Policy and the Economy conference, October 2002.
- Figlio, D. N. og M. Lucas (2000): What's in a grade? School report cards and house prices. *NBER Working Paper Series* 8019.
- Figlio, D. N. og M. Page (2002): Can school choice and school accountability successfully coexist? I C. Hoxby (red): *The economic analysis of school choice*. University of Chicago Press.
- Figlio, D. N. og L. S. Getzler (2002): Accountability, ability and disability: Gaming the system? Mimeo, April 2002.
- Gibbons, R. (1998): Incentives in organizations. *Journal of Economic Perspectives* 12, 115-132
- Jacob, B. A. (2002): Accountability, incentives and behaviour: The impact of high-stakes testing in the Chicago public schools. *NBER Working Paper Series* 8968.
- Jacob, B. A. og S. D. Levitt (2002a): Rotten apples: An investigation of the prevalence and predictors of teacher cheating. Mimeo. University of Chicago, Department of Economics.
- Jacob, B. A. og S. D. Levitt (2002b): Catching cheating teachers: The results of an unusual experiment in implementing theory. *NBER Working Paper Series* 9414.

- Hanushek, E. A. og M. E. Raymond (2002): Improving educational quality: How best to evaluate our schools. Mimeo. Hoover Institution, Stanford University.
- Kane, T. J. og D. O. Staiger (2001): Improving school accountability measures. *NBER Working Paper Series 8156*.
- Ladd, H. F. og R. P. Walsh (2002): Implementing value-added measures of school effectiveness: getting the incentives right. *Economics of Education Review 21*, 1-17.
- Ladd, H. F. og A. Zelli (2001): School-based accountability in North Carolina: The responses of school principals. *Terry Sanford Institute of Policy, Duke University, Working Paper Series 01-13*.
- Lavy, V. (2002a): Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy 110*, 1286-1317.
- Lavy, V. (2002b): Paying for performance: The effect of teacher's financial incentives on students' scholastic outcomes. Mimeo. The Hebrew University of Jerusalem.

