

OM INNRETNINGEN PÅ SKOLEREFORMER

Av Hans Bonesrønning, Torberg Falch og Bjarne Strøm,
Institutt for samfunnsøkonomi, NTNU

Det foregår en omfattende forsøksvirksomhet i grunnskolen for å bedre kvaliteten på undervisningen. Etter vårt syn vil sannsynligvis mange av forsøkene ha liten effekt fordi de ikke påvirker insentivene til aktørene i skolen. Videre vil det være vanskelig å lære noe av forsøksvirksomheten fordi den er svært fragmentert, har uklare målsettinger og mangler tilfredsstillende evalueringsopplegg. Vi argumenterer for at faglig framgang bør være et helt sentralt mål og at forsøk som knytter insentiver til utviklingen i faglig framgang må prioriteres. Slike forsøk bør, i hvert fall i første omgang, knyttes til skolene og ikke til enkeltlærere.

1. Innledning

Det er en allmenn oppfatning at kompetanse og kunnskaper blir stadig viktigere i arbeidslivet og at et godt utdanningssystem er viktig for den økonomiske utviklingen. Dette resonnementet synes å være årsaken til at utdanning har blitt et sentralt politisk tema de siste årene.

Grunnskolen er særlig viktig fordi den legger grunnlaget for videre læring, og fordi den omfatter hele befolkningen. Nye empiriske studier tyder da også på at land der elevene lærer mye i grunnskolen har en sterkere økonomisk vekst enn andre land, se Hanushek og Kimko (2000) og Barro (2001). Det er også flere indikasjoner på at befolkningen er opptatt av skolekvalitet. Blant annet viser undersøkelser av boligprisforskjeller i USA [Black (1999)] og Storbritannia [Gibbons og Machin (2001)] at befolkningen er villig til å betale betydelig mer for boliger i områder med gode skoler. På denne bakgrunn er det viktig at økonomer også deltar i debatten om hvordan skolen bør organiseres.

De siste to årene er det igangsatt en rekke forsøk i den norske grunnskolen. Flere skoler og kommuner har satt i gang forsøksordninger for å forbedre skolekvaliteten. Skoler og kommuner kan søke om fritak for bestemmelser i læreplaner til å gjennomføre ulike forsøk, og søknadene blir som oftest innvilget. Det er stor variasjon i forsøkene, noen innfører fleksitid for elevene, undervisning for flere klasser samtidig der elevene selv bestemmer om de har utbytte av å være tilstede, fast arbeidstid for lærerne, osv. Regjeringen har også

opprettet et 'kvalitetsutvalg' som skal vurdere tiltak for å bedre kvaliteten i skolen. Dette utvalget er ventet å komme med en innstilling i nær framtid.

I dette innlegget vil vi først argumentere for at utviklingen i faglige prestasjoner bør være et hovedelement i vurderingen av effekten av ulike forsøksordninger. Deretter argumenterer vi for at det er behov for endringer i norsk skole. Men siden grunnskolen har vært svært ensrettet i en lang periode, har vi få erfaringer fra Norge om hva som er gunstige systemendringer. Den offentlige debatten preges da også av stor uenighet av effekten av ulike virkemidler.

Endringsviljen synes å være stor, men det er lite kunnskap om hvordan skolepolitikken bør uformes for å skape en bedre skole. Vi diskuterer hvordan vi kan frambringe kunnskap om hvordan ulik skoleorganisering påvirker elevenes læring.. Til slutt gir vi en vurdering av hva vi ser på som fornuftige endringer basert på internasjonale studier, men understreker samtidig at slike endringer bør starte som kontrollerte forsøksordninger.

2. Skolens målsetting og skoleevalueringer

Før man innfører endringer og forsøk i skolen, må det være klart hvordan graden av måloppnåelse skal etterprøves. Et grunnleggende spørsmål i den sammenheng er: Hva er skolens målsettinger? Eksisterende læreplaner inneholder mange generelle og komplekse målsettinger som er lite konkrete. Et problem med *mange* mål er at man mister fokus på de mest grunnleggende målsettingene, noe som kan føre til svikt på de viktigste områdene. Det å ha mange mål blir lett det samme som å ha ingen mål. Et problem med *diffuse* mål er at det er vanskelig å måle i hvilken grad de oppfylles. Kombinasjonen av mange og diffuse målsettinger vil lamme enhver evaluering av forsøksordninger og reformer. Når man ikke er klar på hva man ønsker å oppnå, er det umulig å finne ut om man har oppnådd det man ønsker. Siden effekten av ulike virkemidler som settes inn bør kunne måles og vurderes, er det behov for en klargjøring og konkretisering av de sentrale målsettingene.

Det endelige målet ved utdanning er knyttet til elevenes framtidige funksjonalitet i arbeidslivet og som borgere. I et slikt perspektiv er for eksempel framtidig arbeidsproduktivitet målt ved individuelle lønninger en suksessindikator. Men med en slik målsetting vil det ta svært mange år før dagens skolesystem kan evalueres, og innen den tid er sannsynligvis skolesystemet endret. Det er behov for mer operasjonelle mål. Det synes å være bred enighet om at høyt faglig prestasjonsnivå hos alle elevene er ett sentralt mål. Det er

også klare tegn til at gode lese-, skrive- og regneferdigheter har økende betydning for produktiviteten i arbeidslivet, se for eksempel Murnane, Willett og Levy (1995). Mange andre målsettinger som framholdes som viktig for skolen er ofte indirekte på den måten at de er viktige for den faglige utviklingen.¹ Det bør derfor være helt sentralt å evaluere hvilken effekt skolepolitiske virkemidler har på det faglige prestasjonsnivået hos elevene. Evalueringssystemet bør være utformet slik at den faglige framgangen til elevene kan måles, og ulike skoler må kunne sammenlignes.

Det er en stor utfordring å etablere evalueringssystemer som gjør at *skolenes* bidrag til elevenes faglige fremgang kan identifiseres og isoleres fra andre faktorer som påvirker den faglige utviklingen. Det er viktig å skille mellom skoleeffekter og effekten av for eksempel den sosiale bakgrunnen til elevene. Det eksisterer en internasjonal litteratur som diskuterer hvordan slike analyser bør gjøres og det publiseres empiriske studier av effektene av konkrete utdanningspolitiske virkemidler på elevenes prestasjoner. En oversikt er gitt i Hanushek (2002). Bonesrønning (2002) som analyserer virkninger av klassestørrelse på ungdomsskoleelevers matematikkprestasjoner, er et eksempel på en norsk analyse som følger denne tilnærmingen. Siden det er mange sosiale bakgrunnsfaktorer som er viktig for elevenes kunnskapsnivå, og disse faktorene ofte er vanskelig å måle, velger man gjerne å bruke en såkalt "value-added" metode. Ved "value-added"-analyser sammenlignes det faglige nivået til elevene på et gitt tidspunkt med deres faglige nivå i en tidligere periode. Dermed kontrolleres det for mye av innflytelsen til uobserverbare bakenforliggende faktorer. Elever på gode skoler vil ha en større faglig framgang enn skoler som ikke fungerer godt. Det er dermed ikke automatisk slik at skoler med en stor andel av elevene med en gunstig sosial bakgrunn for læring vil framstå som bedre enn andre skoler. Selv om det kan innvendes mot slike analyser at også veksten i faglig framgang kan avhenge av uobserverbare bakgrunnsfaktorer er "value-added" metoden et betydelig framskritt i forhold til mer naive sammenligninger av resultatnivå på tvers av skoler og gir mindre grad av skjevheter i estimerte skoleeffekter. En kritisk og grundig gjennomgang av skoleevalueringer basert på "value-added" tilnærmingen fins for øvrig i Ladd og Walsh (2002).

¹ I Oslo er det i løpet av våren 2002 utarbeidet et system for evaluering av skolene. Dette inneholder svært mange elementer, se <http://www.skoleetaten.oslo.kommune.no/default.asp?page=/Kvalitet-i-skolen>.

Først når målsetting og evalueringsmetode er klarlagt, er det meningsfullt å gjennomføre forsøksordninger. Dagens situasjon er dessverre ikke slik. Både målsettinger, evalueringsform og innretningen på forsøkene er som oftest svært uklare. Bare det faktum at mange og helt forskjellige forsøk startes samtidig bidrar til at det er svært krevende å etablere generell kunnskap. I tillegg kompliseres dette av at forsøksvirksomheten drives fram av enkeltskoler som i de aller fleste tilfellene ikke evalueres av en utenforstående instans. Intern evaluering kan være verdifull, men det har sine klare begrensninger. For det første er det alltid vanskelig å generalisere på basis av "case"-studier. For det andre er det ofte uklart hva som er sammenligningsgrunnlaget ved interne evalueringer. Dessuten forblir slike evalueringer ofte intern informasjon som ikke tilflyter allmennheten og beslutningstakerne.

3. Behovet for skolereformer

For å gi en vurdering av hvordan den norske skolen fungerer, trenger man et sammenligningsgrunnlag. Det kan i prinsippet være et normativt mål, men det er enklere å sammenligne med et objektive mål. En åpenbar mulighet er å sammenligne prestasjonene til norske elever med prestasjonene til elever i andre land. Siden 1960-tallet er det foretatt internasjonale sammenligninger av faglig dyktighet i grunnskolen. De siste undersøkelsene er TIMSS-studiene i 1995 og 1999 samt PISA-studien i 2000.² Ved alle slike studier legges det stor vekt på å lage sammenlignbare tester i ulike land, og at elevene og skolene utgjør et representativt utvalg i hvert land. TIMSS og PISA har imidlertid en noe ulik innretning. I PISA er det utbyttet av den totale basisutdanningen som er fokusert, mens man i TIMSS har lagt vekt på prestasjoner med utgangspunkt i en felles kjerne av lærerplanelementer i de ulike land. Mens man i TIMSS har som målsetting å måle kunnskaper i forhold til det som undervises, er målsettingen i PISA å måle det som utdanningssystemet bør vektlegge for å forberede elevene for voksenlivet og livslang læring, se Lie m. fl. (2001, kap. 4).

Gjennomsnittresultater for ulike tester for utvalgte land er rapportert i tabell 1. Tallene er konstruert basert på svarene på en rekke oppgaver som er vektet etter vanskelighetsgrad. De er skalert slik at gjennomsnittet over alle land som deltar i en undersøkelse er lik 500 (unntatt

² TIMSS er Third International Mathematics and Science Study utført av International Association for the Evaluation of Educational Achievement (IEA), se www.timss.org. PISA, Programme for International Student Assessment, er et gjennomført av OECD, se www.pisa.oecd.org. Begge disse programmene har planlagt flere oppfølginger av de eksisterende studiene.

TIMSS-studien for 1999). Prestasjonene i et land i forhold til gjennomsnittet er derfor avhengig av hvilke land som deltar i undersøkelsene.³

Tabell 1. Indeks for elevprestasjoner i internasjonale sammenlignbare tester. Noen utvalgte land

Studie	PISA-studien			TIMSS-studien		
	våren 2000			1998/99	1994/95	
	15-åringer			13-åringer	13-åringer	
	Lesing	Matematikk	Naturfag	Matematikk	Matematikk	Naturfag
Finland	546	536	538	520	-	-
Canada	534	533	529	531	527	531
Sør-Korea	525	547	552	587	607	565
Storbritannia	523	529	532	496	506	552
Japan	522	557	550	579	605	571
Sverige	516	510	512	-	519	535
Island	507	514	496	-	487	494
Østerrike	507	515	519	-	539	558
Norge	505	499	500	-	503	527
Frankrike	505	517	500		538	498
USA	504	493	499	502	500	534
Danmark	497	514	481	-	502	478
Spania	493	476	491	-	487	517
Tyskland	484	490	487	-	509	531
Ungarn	480	488	496	532	537	554
Portugal	470	447	459	-	454	480
Latvia	458	463	460	505	493	485
Gjennomsnitt for alle deltakende land	500	500	500	487	500	500

Norske elever presterer forholdsvis nært gjennomsnittet i alle undersøkelsene der Norge har deltatt. For alle testene er prestasjonene på nivå med USA, mens de er gjennomgående svakt

³ I alle studiene er det noen mindre utviklede land som trekker gjennomsnittet ned. I PISA-studien deltok for eksempel Russland, Mexico og Brasil, mens i TIMSS-studien i 1995 deltok blant annet Iran, Colombia og Sør-Afrika. Alle disse landene presterer dårlig.

dårligere enn i Sverige og svakt bedre enn i Danmark. Finske elever presterer svært godt i PISA-studien, mens de Østasiatiske landene gjør det svært godt i realfag.

På grunnlag av de internasjonale testene av kunnskapsnivået til norske elever, kan man i hvert fall si at det er et forbedringspotensial i norsk skole. Siden ressursinnsatsen i norsk skole er høy i forhold til de aller fleste land, blir forbedringspotensialet spesielt synlig dersom man fokuserer på elevprestasjoner per krone. Også i mange andre land er det en utbredt forståelse om at det er et forbedringspotensial i den obligatoriske skolen. Det gjenspeiles i at skolepolitikk er blitt et viktig politisk tema. En del land har innført ulike endringer med sikte på å forbedre skoleprestasjonene. For eksempel har Storbritannia innført større valgfrihet for foreldrene i valg av skole samt prestasjonslønn for lærerne, og Sverige har tillatt private skoler (friskoler) å konkurrere med offentlige skoler.⁴

4. Hvordan bør skolesystemet endres?

For økonomer er det naturlig å vurdere forsøkene i skolen i lys av hvilke virkninger de har på insentivene til aktørene. I dette perspektivet virker mange av forsøkene som gjøres i norsk skole dårlig begrunnet. Mange påvirker overhodet ikke insentivene til aktørene i skolen; skoleledere, lærere, foreldre og elever. For å øke insentivene til gode prestasjoner, må det knyttes en forbindelseslinje mellom oppnådde resultater og belønning.

Folk flest har en klar oppfatning av at lærerne er viktige for elevenes prestasjoner og at elevenes utbytte av undervisningen varierer mye mellom en god og en dårlig lærer, noe som også er påvist i internasjonal utdanningsøkonomisk forskning, se for eksempel Rivkin, Kain og Hanushek (2001). Denne forskningen indikerer at lærerne betyr minst like mye for elevprestasjonene som elevenes familiebakgrunn, men at lærerkvalitet i liten grad henger sammen med lærernes utdanning og erfaring. Det kan derfor synes fornuftig å bruke virkemidler som stimulerer lærerne til økt innsats, kanskje i form av arbeidstid og intensitet, men også i form av søken etter gode undervisningsmetoder.⁵

I det siste har flere har tatt til orde for å øke lønnen til de beste lærerne for å stimulere til økt innsats og gjøre læreryrket mer attraktivt. Et aktuelt eksempel er Oslo, hvor skoledirektøren

⁴ Se www.friskola.se.

⁵ Det norske arbeidsmarkedet for lærere er analysert i Bonesrønning, Falch og Strøm (2001).

har bestemt at rektorene på hver skole skal plukke ut et gitt antall lærere (5 lærere for de fleste skolene) som skal få høyere lønn. Blant lærere og skoleledere er det naturligvis bekymring for hva som *vil* belønnes og uenighet om hva som *bør* belønnes. Stor vektlegging av faglig utvikling tilsier belønning av lærere som bidrar mest til prestasjonsframgang for elevene. For at systemet skal fungere etter hensikten, må skoleledelsen ha insentiver til å belønne akkurat slike lærere og også være i stand til å identifisere disse på formelle eller uformelle måter.

Etter vår vurdering er det et svært omfattende prosjekt å måle hver enkelt lærers betydning for elevens læring på en objektiv måte. For eksempel er familiebakgrunn svært viktig for elevenes prestasjoner. Men lærere som underviser elever med problematisk familiebakgrunn må ikke straffes med fravær av lønnsplag av den grunn. Et individuelt innrettet prestasjonslønnssystem må belønne den innsatsen som enkeltlærerne faktisk gjør, målt som bidrag til økt faglig nivå hos elevene, og ikke spesielle karakteristika ved den elevgruppen som undervises. Lærernes bidrag vil være lettere å identifisere på skolenivå. Analyser på skolenivå har mindre grad av usikkerhet og færre metodiske problemer.

Disse problemene er etter vår vurdering et viktig argument for at økonomiske insentiver bør rettes mot skolene og ikke enkeltlærerne. Det er også opplagt mye mer omfattende å gjøre en evaluering på lærernivå enn å gjøre den på skolenivå. I det minste bør man derfor starte med å evaluere skolene, og deretter eventuelt vurdere om man også ønsker å evaluere enkeltlærerne.

Uansett tror vi at et belønningssystem for enkeltlærere i stor grad må baseres på subjektive vurderinger. Siden det er vanskelig å gi presise kriterier for hva som skal utløse økt lønn, blir det lett misnøye. Evalueringen som ligger til grunn for premiering må oppfattes som rettferdig, og uten klare kriterier kreves det tiltro til den som gjør evalueringen. Lederen som skal ta beslutningen må ha legitimitet, og graden av legitimitet varierer sannsynligvis ganske mye mellom skoler. Differensiering av lønninger etter lærernes evne til å frembringe gode elevprestasjoner vil ofte være ubehagelige beslutninger for rektorene. Dersom rektorenes egen situasjon er uavhengig av om de "riktige" beslutningene tas, er det grunn til å tro at de vil basere seg på relativt etablerte prinsipper som ansiennitet eller utdanningslengde. Derfor kan det som synes som en suksess et sted være helt mislykket et annet sted. Dersom man skal innføre prestasjonslønn, bør det samtidig settes inn virkemidler for å øke skoleledelsens generelle kompetanse.

Hva vet vi om effekten av lignende belønningssystemer fra andre land? Flere delstater i USA har eksperimentert med systemer for prestasjonslønn, men de fleste undersøkelsene derfra finner ingen målbar effekt på elevenes prestasjoner. Samtidig har de aller fleste skoler som har innført prestasjonslønn i hovedsak avsluttet systemet innen 5 år. Et interessant arbeid av Lavy (2002) evaluerer et skoleforsøk i Israel med en annen innretning. Noen skoler ble plukket ut til å delta i et eksperiment der et lønnstillegg til alle lærerne ble gjort avhengig av rankingen av skolene etter forbedring av elevprestasjoner over to år. Undersøkelsen viser at prestasjonsøkningen i de skolene som var med i forsøket var betydelig større enn i skolene som ikke var med. Det indikerer at skolebaserte belønningssystemer som er koblet til målbare prestasjonsendringer hos elevene har positiv effekt på elevenes læringsutbytte via lærernes innsats.

Andre studier konkluderer med at konkurranse fra andre skoler påvirker prestasjonene. Slik konkurranse gir et insitament til å prestere bedre fordi skolene påføres et tap i form av lavere elevtilgang ved lave prestasjoner.. For eksempel finner Hoxby (2000) for USA at de offentlige skolene presterer bedre når det er mange skoledistrikter i et geografisk område, og dermed lave kostnader for foreldrene forbundet med å skifte skoledistrikt. Bergström og Sandström (2001) finner for Sverige at prestasjonene i offentlige skoler øker som følge av at friskoler etableres i nærheten.

Disse studiene er også interessant fordi de illustrerer noen sentrale metodiske poeng når det gjelder utforming og etterprøving av reformforsøk i skolen: For å få kunnskap om hvilke tiltak som virker og ikke virker, er det helt avgjørende at det er mulig å etterprøve effekten på konkrete målsettinger. For å undersøke effekten på elevenes faglige prestasjoner, må det foretas jevnlig tester av elevenes faglige nivå. Disse må suppleres med opplysninger om elevene, lærerne og skolen, og en må kunne foreta sammenligninger mellom skoler som omfattes av et tiltak og skoler som ikke omfattes. I lys av dette framstår dagens forsøksvirksomhet i norsk grunnskole som lite gjennomtenkt. Det gjennomføres en rekke mer eller mindre tilfeldige forsøk og det foreligger, så langt vi har brakt på det rene, en nesten total mangel på systematisk innsamling av informasjon om elevprestasjoner som kan brukes til å evaluere tiltakenes effekt på elevenes læringsutbytte.

5. Avslutning

Vi har argumentert for at skolene bør gis incentiver til å frambringe gode elevprestasjoner . Et problem ved å bruke belønningssystemer rettet mot for eksempel enkeltlærere er at de i stor grad vil bli basert på skjønn og ikke objektive vurderinger. En eventuell objektiv vurdering av enkeltlærere vil være svært omfattende. Hovedproblemene i reformprosessen i norsk skole er at mange tiltak er dårlig begrunnet i eksisterende kunnskap om produksjonsprosessene i klasserommet, og at effekten av mange av tiltakene har vært og er svært vanskelig å etterprøve. Resultatet blir lett at den offentlige debatten blir preget av synspunkter på utformingen av skolepolitikken uten god kunnskap om effekten av ulike tiltak. Parallellen til faget medisin er slående og lite flatterende for skolepolitikken. I medisin er det utenkelig å gjennomføre storstilt bruk av et nytt medikament uten at det er gjennomført kontrollerte eksperimenter med dokumentasjon av virkningene av medikamentet. Selv om de umiddelbare virkninger av ulike medikamenter selvsagt kan være mer dramatiske enn virkninger av skolepolitiske tiltak, er det åpenbart at vi må utforme forsøk i skolen og tilrettelegge for evaluering slik at vi kan sortere ut gode og dårlige forslag til forbedringer.

Referanser:

Barro, R.J. (2001): Human capital and growth. *American Economic Review, Papers and Proceedings* 91, 12-17.

Bergström, F. og M.F. Sandström (2001): Konkurens mellan skolor – för barnens skull! *Ekonomisk Debatt* 29, nr. 5, 319-331.

Bonesrønning, H. (2002): Class size effects on student achievement: Patterns and explanations. Arbeidsnotat, Institutt for samfunnsøkonomi, NTNU.

Bonesrønning, H., T. Falch og B. Strøm (2001): Teacher sorting, teacher quality and student composition. Arbeidsnotat, Institutt for samfunnsøkonomi, NTNU.

Black, S. (1999): Do better schools matter? Parental valuation of elementary Education. *Quarterly Journal of Economics* 114, 577-599.

Gibbons, S. og S. Machin (2001): Valuing primary schools. Working Paper, Center for the Economics of Education. London School of Economics.

Hanushek, E. A. (2002): Publicly provided education. NBER Working Paper 8799.

Hanushek, E. A. og D. D. Kimko (2000): Schooling, labor force quality, and the growth of nations. *American Economic Review* 90, 1184-1208.

Hoxby, C. M. (2000): Does competition among public schools benefit students and taxpayers? *American Economic Review* 90, 1209-1238.

Ladd, H.F. og R.P. Walsh (2002): Implementing value-added measures of school effectiveness: Getting the incentive right. *Economics of Education Review* 21, 1-17.

Lavy, V. (2002): Evaluating the effect of teachers' performance incentives on pupil achievement. Kommer i *Journal of Political Economy*.

Lie, S., M. Kjærnsli, A. Roe og A. Turmo (2001): Godt rustet for framtida? Norske 15-åringers kompetanse i lesing og realfag i et internasjonalt perspektiv. Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Murnane, R.J., J.B. Willett og F. Levy (1995) The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics* 77, 251-66